Title	Finsler interpolation inequalities
Author(s)	Ohta, Shin-ichi
Citation	Calculus of Variations and Partial Differential Equations (2009), 36(2): 211-249
Issue Date	2009-10
URL	http://hdl.handle.net/2433/87313
Right	c Springer-Verlag 2009.
Type	Journal Article
Textversion	publisher

Finsler interpolation inequalities[*][†]

Shin-ichi OHTA[‡]

Department of Mathematics, Faculty of Science, Kyoto University,

Kyoto 606-8502, JAPAN (e-mail: sohta@math.kyoto-u.ac.jp)

Abstract

We extend Cordero-Erausquin, McCann and Schmuckenschläger's Riemannian Borell-Brascamp-Lieb inequality to Finsler manifolds. Among applications, we establish the equivalence between Sturm, Lott and Villani's curvature-dimension condition and a certain lower Ricci curvature bound. We also prove a new volume comparison theorem for Finsler manifolds which is of independent interest.

1 Introduction

Optimal transport theory is making rapid and breathtaking progress in recent years as it finds a large number of connections with various fields. Villani's massive lecture notes [Vi2] provide the global picture, and a great deal of new insight as well. One of the milestones of the theory is a Riemannian Borell-Brascamp-Lieb inequality due to Cordero-Erausquin, McCann and Schmuckenschläger [CMS1] that we will extend to Finsler manifolds. Their work offers deep inspiration as well as a technical breakthrough in the investigation of optimal transport in curved spaces. Furthermore, the Riemannian Borell-Brascamp-Lieb inequality has many meaningful applications from its ancestors (the Prékopa-Leindler inequality and the Brunn-Minkowski inequality) to the attractive curvature-dimension condition. We will also extend them to Finsler manifolds.

Our generalization is twofold. We consider a Finsler manifold where one naturally encounters a nonsymmetric distance function, and we equip it with an arbitrary measure. Then the main difficulty arises from the lack of a good notion of the Hessian, and we always have to take care on the nonsymmetric distance. Nevertheless, surprisingly enough, the conclusion is completely the same as the Riemannian case (see Corollary 8.3 for a variant derived from Theorem 1.2).

Theorem 1.1 (A Finsler Borell-Brascamp-Lieb inequality) *Let (M, F) be a connected, forward geodesically complete, n-dimensional C^∞-Finsler manifold and let m be an arbitrary positive C^∞-measure on M. Take three nonnegative measurable functions f, g, h :*

$M \longrightarrow [0, \infty)$ *and measurable sets* $A, B \subset M$ *with* $\int_A f \, d\mathsf{m} = \int_B g \, d\mathsf{m} = 1$. *If there is* $t \in (0, 1)$ *such that*

$$\frac{1}{h(z)^{1/n}} \leq (1 - t)\left(\frac{\mathfrak{v}_t^>(x, y)}{f(x)}\right)^{1/n} + t\left(\frac{\mathfrak{v}_t^<(x, y)}{g(y)}\right)^{1/n} \tag{1.1}$$

holds for all $x \in A$, $y \in B$ *and* $z \in Z_t(x, y)$, *then we have* $\int_M h \, d\mathsf{m} \geq 1$.

Here $Z_t(A, B)$ denotes the set consisting of points $\gamma(t)$ such that $\gamma : [0, 1] \longrightarrow M$ is a minimal geodesic with $\gamma(0) \in A$ and $\gamma(1) \in B$. In addition, $\mathfrak{v}_t^<$ and $\mathfrak{v}_t^>$ are the *volume distortion coefficients* defined by

$$\mathfrak{v}_t^<(x, y) := \lim_{r \to 0} \frac{\mathsf{m}(Z_t(x, B^+(y, r)))}{\mathsf{m}(B^+(y, tr))}, \quad \mathfrak{v}_t^>(x, y) := \lim_{r \to 0} \frac{\mathsf{m}(Z_t(B^-(x, r), y))}{\mathsf{m}(B^-(x, (1 - t)r))} \tag{1.2}$$

($<$ and $>$ represent shapes of cones we have in mind), where $B^+(x, r)$ and $B^-(x, r)$ are forward and backward open balls with center x and radius r (see (2.4)).

In order to obtain more concrete statements, we need to control the volume distortion coefficients by means of geometric quantities of (M, F, m). Shen's volume comparison theorem ([Sh1], Theorem 7.1, see also [Sh2] for applications) is usable, whereas we also give a new, sharper comparison theorem (Theorem 7.3) inspired by the theory of weighted Riemannian manifolds. This theorem would be even more significant than Theorem 1.1 from the geometric viewpoint and there are a lot of analytic and geometric applications (see [OS] for recent progress), though we do not pursue that direction in this article.

One of the most sophisticated applications of the technique developed in the proof of Theorems 1.1 and 7.3 is Theorem 1.2 below in which we successfully generalize the curvature-dimension condition (also called the N-Ricci curvature bound) recently remarkably developed by Sturm [St2], [St3] and Lott and Villani [LV1], [LV2]. Again it is surprising that the statement is completely similar to the weighted Riemannian situation.

Theorem 1.2 (The curvature-dimension condition) *Let* (M, F, m) *be as in Theorem* 1.1 *with* $n \geq 2$ *and take* $K \in \mathbb{R}$.

(i) *For* $N \in (n, \infty)$, (M, F, m) *has N-Ricci curvature bounded below by K if and only if we have*

$$\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} - \frac{1}{N - n}(\partial_v \mathcal{V})^2 \geq K \tag{1.3}$$

for every unit vector $v \in TM$.

(ii) (M, F, m) *has n-Ricci curvature bounded below by K if and only if* $\mathrm{Ric}(v) \geq K$ *and* $\partial_v \mathcal{V} = 0$ *hold for every unit vector* $v \in TM$.

(iii) (M, F, m) *has ∞-Ricci curvature bounded below by K if and only if* $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} \geq K$ *holds for every unit vector* $v \in TM$.

Here we set, given a unit vector $v \in T_x M$,

$$\mathcal{V}(v) := \log \left(\frac{\mathrm{vol}_{g_v}(B^+_{T_x M}(0,1))}{\mathsf{m}_x(B^+_{T_x M}(0,1))} \right), \tag{1.4}$$

where vol_{g_v} and m_x stand for Lebesgue measures on $T_x M$ induced from g_v and m, respectively (see (2.2) for the definition of g_v). We can rewrite this as $\mathsf{m}_x(B^+_{T_x M}(0,1)) = e^{-\mathcal{V}(v)} \mathrm{vol}_{g_v}(B^+_{T_x M}(0,1))$. Then we define

$$\partial_v \mathcal{V} := \frac{d}{dt}\Big|_{t=0} \mathcal{V}\big(\dot{\eta}(t)\big), \qquad \partial^2_v \mathcal{V} := \frac{d^2}{dt^2}\Big|_{t=0} \mathcal{V}\big(\dot{\eta}(t)\big), \tag{1.5}$$

where $\eta : (-\varepsilon, \varepsilon) \longrightarrow M$ is the geodesic with $\dot{\eta}(0) = v$. We remark that constant multiplication $C\mathsf{m}$ does not affect $\partial_v \mathcal{V}$ nor $\partial^2_v \mathcal{V}$, it just turns $\mathcal{V}$ into $\mathcal{V} - \log C$. Theorem 1.2 guarantees that our curvature bound (with weight) is not artificial, for the curvature-dimension condition involves only the distance and the measure (see Definition 8.1).

It is known that Banach spaces have nonnegative Ricci curvature in terms of the curvature-dimension condition (due to Cordero-Erausquin, see [Vi2, Theorem in page 908]). This means that the curvature-dimension condition does not characterize Riemannian spaces. From a different viewpoint, the curvature-dimension condition can be a powerful tool also in the investigation of Finsler spaces. This is what we have done in this article. In fact, some of standard applications of our theorems, such as the Lichnerowicz inequality (Corollary 8.5) and normal concentration of measures (Proposition 9.6), seem new for general Finsler manifolds.

We explain an outline of the proof of Theorem 1.1. We first establish a Finsler analogue of the Brenier-McCann solution to the Monge-Kantorovich problem (Theorem 4.10). That is to say, a unique optimal way of transporting one probability measure to another is doing along the gradient vector field $\nabla(-\phi)$ for some c-concave function ϕ, where $c = d^2/2$. Although we will give a self-contained proof along the same lines as McCann's [Mc], this part follows also from Villani's [Vi2, Chapter 10] (see also [BB1], [BB2], [FF]). He treats quite general Lagrangian cost functions on Riemannian manifolds, while the underlying Riemannian structure plays only a subsidiary role.

We next analyze the c-concave function ϕ to prove Theorem 1.1. The point is to give up trying to formulate a Finsler Hessian and use the second order differential $d(d\phi)$: $TM \longrightarrow T(T^*M)$. Even more important from the technical viewpoint is to take out vertical terms of $d(d\phi)$ (see Lemma 3.2, Proposition 5.1). Our argument makes it clear what the Hessian is implicitly doing in [CMS1]. Once we realize this method, it turns out that we can closely follow the proof in [CMS1]. Moreover, by virtue of our arbitrarily fixed underlying measure m, Theorem 1.1 includes a weighted version studied in [CMS2].

The article is organized as follows. We start with a review of Finsler geometry in Section 2, and show technical lemmas in Section 3. Then we study the Brenier-McCann solution to the Monge-Kantorovich problem in Section 4. Sections 5 and 6 are devoted to the proof of Theorem 1.1. A volume comparison theorem is established in Section 7. We discuss the curvature-dimension condition in Section 8. We finally present selected applications in Section 9.

Throughout the article, unless otherwise indicated, (M, F) is a connected, forward geodesically complete, n-dimensional C^∞-Finsler manifold, $c(x,y) := d(x,y)^2/2$ is the

quadratic cost function and m is an arbitrary positive C^∞-measure on M. We emphasis that F is merely positively homogeneous, therefore c is generally nonsymmetric. Every geodesic will have a constant speed.

2 Preliminaries for Finsler geometry

In this section, we review the fundamentals of Finsler geometry. Basic references are [BCS] and [Sh2].

2.1 Finsler structures and the Legendre transform

Let M be a connected, n-dimensional C^∞-manifold. For $x \in M$, denote by $T_x M$ the tangent space at x, put $TM := \bigcup_{x \in M} T_x M$ and let $\pi : TM \longrightarrow M$ be the natural projection. Given a local coordinate system $(x^i)_{i=1}^n : U \longrightarrow \mathbb{R}^n$ on an open set $U \subset M$, we will always denote by $(x^i, v^i)_{i=1}^n$ the local coordinate system on $\pi^{-1}(U) \subset TM$ given by, for $v \in \pi^{-1}(U)$,

$$v = \sum_{i=1}^n v^i \frac{\partial}{\partial x^i}\Big|_{\pi(v)}.$$

Definition 2.1 (Finsler structures) A C^∞-*Finsler structure* of a C^∞-manifold M is a function $F : TM \longrightarrow [0, \infty)$ satisfying the following conditions:

(1) (Regularity) The function F is C^∞ on $TM \setminus 0$, where 0 stands for the zero section.

(2) (Positive homogeneity of degree 1) For any $v \in TM$ and positive number $\lambda > 0$, we have $F(\lambda v) = \lambda F(v)$.

(3) (Strong convexity) Given a local coordinate system $(x^i)_{i=1}^n$ on $U \subset M$, the $n \times n$ matrix
$$\big(g_{ij}(v)\big) := \left(\frac{1}{2} \frac{\partial^2 (F^2)}{\partial v^i \partial v^j}(v) \right)$$
is positive-definite at every $v \in \pi^{-1}(U) \setminus 0$.

In other words, each tangent space $(T_x M, F)$ is a *Minkowski space* and F varies C^∞ in the horizontal direction. The above definition has collected as few conditions as possible, other expected properties follow from them (e.g., $F > 0$ on $TM \setminus 0$, strict convexity of F etc.). We emphasis that, however, F is not necessarily *absolutely homogeneous*, namely $F(v) \neq F(-v)$ may happen. It is sometimes helpful to consider the *reverse* of F, $\bar{F}(v) := F(-v)$, which turns everything around (e.g., distance and geodesics).

The *Legendre transform* $\mathfrak{L} : T^*M \longrightarrow TM$ associates each co-vector $\alpha \in T_x^*M$ with a unique vector $v = \mathfrak{L}_x(\alpha) \in T_x M$ such that $F(v) = F^*(\alpha)$ and $\alpha(v) = F^*(\alpha)^2$, where F^* denotes the dual Minkowski norm of F on T_x^*M. (This $\mathfrak{L}$ is the inverse of that in [BCS,

§14.8].) The transform $\mathfrak{L}$ is a C^∞-diffeomorphism from $T^*M \setminus 0$ to $TM \setminus 0$, and C^0 on T^*M as well. For a C^1-function $f : M \longrightarrow \mathbb{R}$, we define the *gradient vector* of f at $x \in M$ as the Legendre transform of its differential, i.e., $\nabla f(x) := \mathfrak{L}_x(df_x) \in T_xM$. Note that, for a unit speed C^1-curve $\eta : [0,l] \longrightarrow M$ (i.e., $F(d\eta/dt) \equiv 1$), it holds that

$$-\int_0^l F\Big(\nabla(-f)\big(\eta(t)\big)\Big)\, dt \leq f\big(\eta(l)\big) - f\big(\eta(0)\big) \leq \int_0^l F\Big(\nabla f\big(\eta(t)\big)\Big)\, dt.$$

Let $f : M \longrightarrow \mathbb{R}$ be a C^2-function. We do not have a good notion of a Finsler Hessian (it is usually just the second order differentiation along geodesics), so that we will use the differentiation of $df : M \longrightarrow T^*M$ instead, namely $d(df)_x : T_xM \longrightarrow T_{df_x}(T^*M)$. In a local coordinate system, it is written as

$$d(df)_x = \sum_{i,j=1}^n \left(\delta^i{}_j \frac{\partial}{\partial x^i}\Big|_{df_x} + \frac{\partial^2 f}{\partial x^i \partial x^j}(x)\frac{\partial}{\partial v^i}\Big|_{df_x}\right) dx^j|_x. \tag{2.1}$$

We remark that the vertical part

$$\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x^i \partial x^j}(x)\frac{\partial}{\partial v^i}\Big|_{df_x} dx^j|_x$$

is coordinate-free only when $df_x = 0$. Note also that the coordinate transform on $T_{df_x}(T^*M)$ between (x^i, v^i) and (y^j, w^j) is

$$\frac{\partial}{\partial x^i} = \sum_{j=1}^n \left\{ \frac{\partial y^j}{\partial x^i}\frac{\partial}{\partial y^j} + \sum_{k=1}^n \frac{\partial f}{\partial x^k}\frac{\partial^2 x^k}{\partial x^i \partial y^j}\frac{\partial}{\partial w^j}\right\}, \quad \frac{\partial}{\partial v^i} = \sum_{j=1}^n \frac{\partial x^i}{\partial y^j}\frac{\partial}{\partial w^j}.$$

Therefore the vertical part behaves well and can be identified with $T_{df_x}(T_x^*M)$, whereas the horizontal part is cumbersome. These observations underlie Lemma 3.2 and Proposition 5.1.

2.2 The Chern connection and covariant derivatives

Given $v \in T_xM \setminus 0$, the $n \times n$ matrix $(g_{ij}(v))$ in Definition 2.1 defines a useful Riemannian structure of T_xM and we denote it by g_v, that is,

$$g_v\left(\sum_{i=1}^n w_1^i \frac{\partial}{\partial x^i}\Big|_x, \sum_{j=1}^n w_2^j \frac{\partial}{\partial x^j}\Big|_x\right) := \sum_{i,j=1}^n g_{ij}(v)w_1^i w_2^j. \tag{2.2}$$

Note that $g_v(v,v) = F(v)^2$ follows from Euler's theorem. Moreover, g_v is linked to the Legendre transform through the formula $\mathfrak{L}_x^{-1}(v)(w) = g_v(v,w)$ for $w \in T_xM$. We call $g_{ij}(v)$ the *fundamental tensor* and further define the *Cartan tensor* by, for $v \in TM \setminus 0$,

$$A_{ijk}(v) := \frac{F(v)}{2}\frac{\partial g_{ij}}{\partial v^k}(v) = \frac{F(v)}{4}\frac{\partial^3 (F^2)}{\partial v^i \partial v^j \partial v^k}(v).$$

If F is coming from a Riemannian structure, then g_v coincides with the original Riemannian structure for all $v \in TM \setminus 0$, and hence the Cartan tensors vanish everywhere. In fact, the converse is also true, therefore F is Riemannian if and only if the Cartan tensors vanish everywhere on $TM \setminus 0$.

We define the *formal Christoffel symbol* by

$$\gamma^i{}_{jk}(v) := \frac{1}{2} \sum_{l=1}^{n} g^{il}(v) \left\{ \frac{\partial g_{lj}}{\partial x^k}(v) - \frac{\partial g_{jk}}{\partial x^l}(v) + \frac{\partial g_{kl}}{\partial x^j}(v) \right\}$$

for $v \in TM \setminus 0$, and also define

$$N^i{}_j(v) := \sum_{k=1}^{n} \gamma^i{}_{jk}(v) v^k - \frac{1}{F(v)} \sum_{k,l,m=1}^{n} A^i{}_{jk}(v) \gamma^k{}_{lm}(v) v^l v^m,$$

where (g^{ij}) stands for the inverse matrix of (g_{ij}) and $A^i{}_{jk} := \sum_l g^{il} A_{ljk}$.

Given a connection ∇ on the pulled-back tangent bundle π^*TM, we denote its connection one-forms by $\omega_j{}^i$, that is,

$$\nabla_v \frac{\partial}{\partial x^j} = \sum_{i=1}^{n} \omega_j{}^i(v) \frac{\partial}{\partial x^i}, \qquad \nabla_v dx^i = -\sum_{j=1}^{n} \omega_j{}^i(v)\, dx^j.$$

Different from the Riemannian situation, there are several connections (due to Cartan, Chern, Berwald and so on) each of which is canonical in its own way. We use only one of them in this article.

Definition 2.2 (The Chern connection) Let (M, F) be a C^∞-Finsler manifold. Then there exists a unique connection ∇ on the pulled-back tangent bundle π^*TM, called the *Chern connection*, whose connection one-forms $\omega_j{}^i$ satisfy the following conditions:

Fix a local coordinate system $(x^i)_{i=1}^n$ on $U \subset M$.

(1) (Torsion-freeness) For any $i = 1, 2, \ldots, n$, we have

$$\sum_{j=1}^{n} dx^j \wedge \omega_j{}^i = 0.$$

(2) (Almost g-compatibility) For any $i, j = 1, 2, \ldots, n$, we have

$$dg_{ij} - \sum_{k=1}^{n} (g_{kj} \omega_i{}^k + g_{ik} \omega_j{}^k) = \frac{2}{F} \sum_{k=1}^{n} A_{ijk}\, \delta v^k,$$

where we set $\delta v^k := dv^k + \sum_l N^k{}_l\, dx^l$.

In the remainder of this section, ∇ always stands for the Chern connection on π^*TM. The torsion-freeness says that the connection one-form $\omega_j{}^i$ does not have any dv^k-term,

so that we can write $\omega_j{}^i = \sum_k \Gamma^i{}_{jk} dx^k$. Together with the almost g-compatibility, we find the explicit formula

$$\Gamma^i{}_{jk} = \gamma^i{}_{jk} - \frac{1}{F} \sum_{l,m=1}^{n} g^{il}(A_{ljm}N^m{}_k - A_{jkm}N^m{}_l + A_{klm}N^m{}_j).$$

If (M, F) is Riemannian, then the almost g-compatibility reduces to the *g-compatibility* $dg_{ij} = \sum_k (g_{kj}\omega_i{}^k + g_{ik}\omega_j{}^k)$, therefore the Chern connection is nothing but the Levi-Civita connection. We say that a Finsler manifold (M, F) is *Berwald type* if $\Gamma^i{}_{jk}(v)$ depends only on $x = \pi(v)$ (i.e., $\Gamma^i{}_{jk}$ is fiber-wise constant). For instance, Riemannian manifolds and Minkowski spaces are Berwald type. Roughly speaking, a Finsler manifold of Berwald type is modeled on a single Minkowski space. Finsler manifolds of Berwald type have already provided a rich family of non-Riemannian spaces.

For a C^∞-vector field X on M and two nonzero vectors $v, w \in T_xM \setminus 0$, we define the *covariant derivative* $D_v^w X$ with *reference vector* w as

$$(D_v^w X)(x) := \sum_{i,j=1}^{n} \left\{ v^j \frac{\partial X^i}{\partial x^j}(x) + \sum_{k=1}^{n} \Gamma^i{}_{jk}(w)v^j X^k(x) \right\} \frac{\partial}{\partial x^i}\Big|_x,$$

where $X(x) = \sum_i X^i(x)(\partial/\partial x^i)|_x$. We usually choose $w = v$ or $w = X(x)$.

2.3 Flag and Ricci curvatures

The Chern connection ∇ derives the corresponding curvature two-form

$$\Omega_j{}^i(v) := d\omega_j{}^i - \sum_{k=1}^{n} \omega_j{}^k \wedge \omega_k{}^i,$$

where we put $\omega \wedge \tau := \omega \otimes \tau - \tau \otimes \omega$. It can be rewritten as

$$\Omega_j{}^i(v) = \frac{1}{2} \sum_{k,l=1}^{n} R_j{}^i{}_{kl}(v) dx^k \wedge dx^l + \frac{1}{F(v)} \sum_{k,l=1}^{n} P_j{}^i{}_{kl}(v) dx^k \wedge \delta v^l,$$

where we impose $R_j{}^i{}_{kl} = -R_j{}^i{}_{lk}$.

Given two linearly independent vectors $v, w \in T_xM \setminus 0$, we define the *flag curvature* by

$$\mathcal{K}(v, w) := \frac{g_v(R^v(w, v)v, w)}{g_v(v, v)g_v(w, w) - g_v(v, w)^2},$$

where we set, for $v = \sum_i v^i(\partial/\partial x^i)|_x$ and $w = \sum_i w^i(\partial/\partial x^i)|_x$,

$$R^v(w, v)v := \sum_{i,j,k,l=1}^{n} v^j R_j{}^i{}_{kl}(v)w^k v^l \frac{\partial}{\partial x^i}\Big|_x.$$

See [Sh2, §6] for a nice geometric interpretation of $\mathcal{K}$. For later convenience, we recall from [BCS, (3.4.7), Exercise 3.4.4(c)] that, given $v \in T_xM \setminus 0$ and $w, w' \in T_xM$,

$$g_v\big(R^v(w, v)v, w'\big) = g_v\big(R^v(w', v)v, w\big), \qquad R^v(v, v)v = 0. \tag{2.3}$$

Unlike the Riemannian case, the flag curvature $\mathcal{K}(v, w)$ depends not only on the *flag* $\{\lambda v + \mu w \,|\, \lambda, \mu \in \mathbb{R}\}$, but also on the *flag pole* $\{\lambda v \,|\, \lambda > 0\}$. One merit of the flag curvature is its independence from the choice of connections.

For $v \in T_x M \setminus 0$, the *Ricci curvature* of v is defined by

$$\mathrm{Ric}(v) := \sum_{i=1}^{n-1} \mathcal{K}(v, e_i),$$

where $e_1, e_2, \ldots, e_{n-1}, v/F(v)$ form an orthonormal basis of $T_x M$ with respect to g_v.

2.4   Geodesics, the exponential map and cut loci

For a C^1-curve $\eta : [0, r] \longrightarrow M$, we define its arclength in a natural way by

$$L(\eta) := \int_0^r F\big(\dot{\eta}(t)\big) \, dt, \qquad \dot{\eta}(t) := \frac{d\eta}{dt}(t).$$

Then the corresponding distance function $d : M \times M \longrightarrow [0, \infty)$ is given by $d(x, y) := \inf_\eta L(\eta)$, where the infimum is taken over all C^1-curves η from x to y. We emphasis that d is not necessarily symmetric (i.e., $d(x, y) \neq d(y, x)$ may happen) because F is merely positively homogeneous. Nonetheless, d is positive outside the diagonal set and satisfies the triangle inequality $d(x, z) \leq d(x, y) + d(y, z)$. We define the *forward* and *backward open balls* of center $x \in M$ and radius $r > 0$ by

$$B^+(x, r) := \{y \in M \,|\, d(x, y) < r\}, \quad B^-(x, r) := \{y \in M \,|\, d(y, x) < r\}. \tag{2.4}$$

We also define open balls in $T_x M$ by

$$B_{T_x M}^+(0, r) := \{v \in T_x M \,|\, F(v) < r\}, \quad B_{T_x M}^-(0, r) := \{v \in T_x M \,|\, F(-v) < r\}.$$

A C^∞-curve $\eta : [0, r] \longrightarrow M$ is called a *geodesic* (of constant speed) if it satisfies $D_{\dot{\eta}}^{\dot{\eta}} \dot{\eta} = 0$ on $(0, r)$. We remark that the reverse curve $\bar{\eta}(t) := \eta(r - t)$ may not be a geodesic ($\bar{\eta}$ is a geodesic with respect to the reverse Finsler structure $\bar{F}$). For C^∞-vector fields V, W along a nonconstant geodesic $\eta : [0, r] \longrightarrow M$, it holds that

$$\frac{d}{dt}\big[g_{\dot{\eta}(t)}(V, W)\big] = g_{\dot{\eta}(t)}(D_{\dot{\eta}}^{\dot{\eta}} V, W) + g_{\dot{\eta}(t)}(V, D_{\dot{\eta}}^{\dot{\eta}} W). \tag{2.5}$$

We define the *exponential map* by $\exp v = \exp_{\pi(v)} v := \eta(1)$ for $v \in TM$ if there is a geodesic $\eta : [0, 1] \longrightarrow M$ with $\dot{\eta}(0) = v$. The exponential map is only C^1 at the zero section, and is C^2 at the zero section if and only if (M, F) is Berwald type. Moreover, the squared distance function $d(x, \cdot)^2$ from a point $x \in M$ is C^2 at x for all $x \in M$ if and only if (M, F) is Riemannian ([Sh1, Proposition 2.2]). The lack of C^2-smoothness is troublesome and we need some extra discussions in later sections which are unnecessary in the Riemannian case.

A Finsler manifold (M, F) is said to be *forward geodesically complete* if the exponential map is defined on the entire TM, in other words, if there is a geodesic $\eta : [0, \infty) \longrightarrow M$

with $\dot{\eta}(0) = v$ for every $v \in TM$. Then any two points in M can be connected by a minimal geodesic, i.e., a geodesic whose arclength coincides with the distance from the initial point to the terminal point.

Fix $x \in M$. For a unit vector $v \in T_x M$, let $r(v) \in (0, \infty]$ be the supremum of $r > 0$ such that the geodesic $t \longmapsto \exp_x tv$ is minimal on $[0, r]$. If $r(v) < \infty$, then $\exp_x(r(v)v)$ is called a *cut point* of x, and the *cut locus* $\mathrm{Cut}(x)$ of x is defined as the set of all cut points of x. The exponential map $\exp_x$ is a C^∞-diffeomorphism from $\{tv \mid v \in T_x M, F(v) = 1, t \in (0, r(v))\}$ to $M \setminus (\mathrm{Cut}(x) \cup \{x\})$. We remark that y is a cut point of x with respect to F if and only if x is a cut point of y with respect to $\bar{F}$.

2.5   Jacobi fields and variational formulas for arclength

A C^∞-vector field J along a geodesic $\eta : [0, r] \longrightarrow M$ is called a *Jacobi field* if it satisfies

$$D_{\dot{\eta}}^{\dot{\eta}} D_{\dot{\eta}}^{\dot{\eta}} J + R^{\dot{\eta}}(J, \dot{\eta})\dot{\eta} = 0$$

on $[0, r]$. Any Jacobi field is represented as the variational vector field of a geodesic variation and vice versa. For C^∞-vector fields V, W along a nonconstant geodesic $\eta : [0, r] \longrightarrow M$, we define the *index form* by

$$I(V, W) := \frac{1}{F(\dot{\eta})} \int_0^r \left\{ g_{\dot{\eta}(t)}(D_{\dot{\eta}}^{\dot{\eta}} V, D_{\dot{\eta}}^{\dot{\eta}} W) - g_{\dot{\eta}(t)}\left(R^{\dot{\eta}}(V, \dot{\eta})\dot{\eta}, W\right) \right\} dt.$$

Note that $I(V, W) = I(W, V)$ holds by (2.3). We deduce from (2.5) that, if V is a Jacobi field,

$$I(V, W) = F(\dot{\eta})^{-1} \left[g_{\dot{\eta}(t)}(D_{\dot{\eta}}^{\dot{\eta}} V, W) \right]_{t=0}^r. \tag{2.6}$$

We consider a C^∞-variation $\sigma : [0, r] \times (-\varepsilon, \varepsilon) \longrightarrow M$ and set

$$T(t, s) := \partial_t \sigma(t, s) = \frac{\partial \sigma}{\partial t}(t, s), \qquad U(t, s) := \partial_s \sigma(t, s) = \frac{\partial \sigma}{\partial s}(t, s).$$

Then the first variation of arclength is written as

$$\frac{\partial L(\sigma_s)}{\partial s} = \left[\frac{g_{T(t,s)}(U, T)}{F(T(t, s))} \right]_{t=0}^r - \int_0^r g_{T(t,s)}\left(U, D_T^T\left[\frac{T}{F(T)} \right] \right) dt, \tag{2.7}$$

where we put $\sigma_s(t) := \sigma(t, s)$. For a C^∞-variation $\sigma : [0, r] \times (-\varepsilon, \varepsilon) \longrightarrow M$ such that σ_0 is a geodesic, its second variation of arclength is written by using the index form as

$$\frac{\partial^2 L(\sigma_s)}{\partial s^2}\bigg|_{s=0} = I(U, U) + \left[\frac{g_{T(t)}(D_U^T U, T)}{F(T(t))} \right]_{t=0}^r - \int_0^r \frac{1}{F(T(t))} \left\{ \frac{\partial F(\partial_t \sigma)}{\partial s}(t) \right\}^2 dt. \tag{2.8}$$

We omitted $s = 0$ in the right-hand side for brevity.

2.6 A word for the underlying measure m

Different from Riemannian manifolds, there is a variety of canonical measures on Finsler manifolds, such as the Busemann-Hausdorff measure, the Holmes-Thompson measure and so forth (see [AT]). On the other hand, as we have weighted measures in our sights, we shall eventually treat all measures. Therefore it is natural to consider an arbitrarily chosen positive C^∞-measure on M in the first place, and we will denote it by m. (To be precise, in any local coordinate system $(x^i)_{i=1}^n$, we can write $\mathsf{m} = \varphi \, dx^1 \cdots dx^n$ using some positive C^∞-function φ. Then we have $\mathsf{m}_x = \varphi(x) \, dv^1 \cdots dv^n$ on T_xM.) See Section 7 for a further convincing discussion.

An advantage of this presentation is that, as we mentioned above, it is a priori covering the weighted version. Take a C^∞-function $V : M \longrightarrow \mathbb{R}$ and consider $\hat{\mathsf{m}} = e^{-V}\mathsf{m}$. For f, g and h as in Theorem 1.1, we set $\hat{f} = fe^V$, $\hat{g} = ge^V$ and $\hat{h} = he^V$. Note that $\int_A \hat{f} \, d\hat{\mathsf{m}} = \int_B \hat{g} \, d\hat{\mathsf{m}} = 1$. Then the corresponding volume distortion coefficients of $\hat{\mathsf{m}}$ are

$$\hat{\mathfrak{v}}_t^<(x,y) = e^{V(y)-V(z)}\mathfrak{v}_t^<(x,y), \qquad \hat{\mathfrak{v}}_t^>(x,y) = e^{V(x)-V(z)}\mathfrak{v}_t^>(x,y)$$

if $Z_t(x,y)$ consists of a single point z (this situation will turn out essential). Thus the hypothesis (1.1) is rewritten as

$$\frac{1}{\hat{h}(z)^{1/n}} \leq (1-t)\left(\frac{\hat{\mathfrak{v}}_t^>(x,y)}{\hat{f}(x)}\right)^{1/n} + t\left(\frac{\hat{\mathfrak{v}}_t^<(x,y)}{\hat{g}(y)}\right)^{1/n},$$

and the conclusion is $\int_M \hat{h} \, d\hat{\mathsf{m}} = \int_M h \, d\mathsf{m} \geq 1$.

3 Technical ingredients

Before beginning to study optimal transport, we prove two rather technical lemmas for later use. See [CMS1] for the Riemannian case.

3.1 A characterization of cut loci

We will use the following lemma in the proof of Proposition 5.1.

Lemma 3.1 *If y is a cut point of x, then the function $f(z) := d(z,y)^2/2$ satisfies*

$$\liminf_{v \to 0 \in T_xM} \frac{f(\xi_v(1)) + f(\xi_v(-1)) - 2f(x)}{F(v)^2} = -\infty,$$

where $\xi_v : [-1,1] \longrightarrow M$ is the geodesic with $\dot{\xi}_v(0) = v$.

Proof. First of all, y is a cut point of x if either there are two minimal geodesics from x to y, or y is the first conjugate point of x along a unique minimal geodesic η from x to y, namely there is a Jacobi field J along η vanishing only at x and y (cf. [BCS, Corollary 8.2.2]).

We first assume that there are two distinct minimal geodesics $\eta, \zeta : [0, d(x,y)] \longrightarrow M$ from x to y. Put $v = \dot\zeta(0)$, $w = \dot\eta(0)$ and $y_\varepsilon = \eta(d(x,y) - \varepsilon)$ for fixed small $\varepsilon > 0$, and note that $y_\varepsilon \notin \mathrm{Cut}(x)$. Then the first variation formula (2.7) shows that, for $t > 0$,

$$f\big(\xi_v(-t)\big) - f(x) \leq \big\{ d\big(\xi_v(-t), y_\varepsilon\big) + \varepsilon \big\}^2 / 2 - \big\{ d(x, y_\varepsilon) + \varepsilon \big\}^2 / 2$$
$$= t\{d(x, y_\varepsilon) + \varepsilon\} g_{\dot\eta(0)}\big(v, \dot\eta(0)\big) + O(t^2)$$
$$= d(x,y) g_w(tv, w) + O(t^2).$$

Here the error term $O(t^2)$ is ensured by the smoothness of the geodesic ξ_v. However, we find

$$f\big(\xi_v(t)\big) - f(x) = \{d(x,y) - t\}^2 / 2 - d(x,y)^2 / 2 = -td(x,y) + t^2 / 2.$$

Therefore we have

$$\frac{f(\xi_v(t)) + f(\xi_v(-t)) - 2f(x)}{t^2} \leq -\frac{1 - g_w(v, w)}{t} d(x,y) + t^{-2} O(t^2) \to -\infty$$

as t tends to zero, for $g_w(v, w) = \mathfrak{L}_x^{-1}(w)(v) < 1$.

We next treat the case where y is the first conjugate point along a unique minimal geodesic $\eta : [0,1] \longrightarrow M$ from x to y. Take a Jacobi field J along η vanishing only at 0 and 1. Put $v = D_{\dot\eta}^{\dot\eta} J(0) \in T_x M \setminus 0$ and let V_1 be the parallel vector field along η (i.e., $D_{\dot\eta}^{\dot\eta} V_1 \equiv 0$) with $V_1(0) = v$. We further define $V(t) := (1-t)V_1(t)$ for $t \in [0,1]$ and $J_\varepsilon := J + \varepsilon V$ for small $\varepsilon > 0$. Note that $J_\varepsilon(0) = \varepsilon v$ and $J_\varepsilon(1) = 0$. In addition, $d[g_{\dot\eta(t)}(J, V_1)]/dt|_{t=0} = g_{\dot\eta(0)}(v, v) > 0$ ensures that $J_\varepsilon \neq 0$ on $[0, 1)$ for sufficiently small $\varepsilon > 0$.

Consider the variation $\sigma : [0,1] \times [-1,1] \longrightarrow M$ given by $\sigma(t,s) = \sigma_s(t) := \xi_{J_\varepsilon(t)}(s)$. We remark that σ is C^∞ on $(0,1) \times (-1,1)$ since $J_\varepsilon \neq 0$ on $[0,1)$. Then the second variation formula (2.8) shows that

$$\left. \frac{\partial^2 L(\sigma_s)}{\partial s^2} \right|_{s=0} = I(J_\varepsilon, J_\varepsilon) - \frac{g_{\dot\eta(0)}(D_{J_\varepsilon}^{\dot\eta} J_\varepsilon, \dot\eta)}{d(x,y)} - \frac{1}{d(x,y)} \int_0^1 \left\{ \frac{\partial F(\partial_t \sigma)}{\partial s}(t) \right\}^2 dt.$$

We put

$$\mathcal{T}_{\dot\eta(0)}(v) := g_{\dot\eta(0)}(D_v^v v - D_v^{\dot\eta} v, \dot\eta) = \varepsilon^{-2} g_{\dot\eta(0)}(D_{J_\varepsilon}^{J_\varepsilon} J_\varepsilon - D_{J_\varepsilon}^{\dot\eta} J_\varepsilon, \dot\eta) = -\varepsilon^{-2} g_{\dot\eta(0)}(D_{J_\varepsilon}^{\dot\eta} J_\varepsilon, \dot\eta)$$

(this quantity is called the *tangent curvature*). The last equality follows from the construction that $\sigma(0, \cdot) = \xi_{J_\varepsilon(0)}$ is a geodesic. These together yield, using (2.6),

$$\left. \frac{\partial^2 L(\sigma_s)}{\partial s^2} \right|_{s=0} \leq I(J, J) + 2\varepsilon I(J, V) + \varepsilon^2 I(V, V) + \varepsilon^2 \mathcal{T}_{\dot\eta(0)}(v)/d(x,y)$$
$$= \left\{ \big[g_{\dot\eta}(D_{\dot\eta}^{\dot\eta} J, J) \big]_{t=0}^1 + 2\varepsilon \big[g_{\dot\eta}(D_{\dot\eta}^{\dot\eta} J, V) \big]_{t=0}^1 + \varepsilon^2 \mathcal{T}_{\dot\eta(0)}(v) \right\}/d(x,y) + \varepsilon^2 I(V, V)$$
$$= \left\{ -2\varepsilon g_{\dot\eta(0)}(v, v) + \varepsilon^2 \mathcal{T}_{\dot\eta(0)}(v) \right\}/d(x,y) + \varepsilon^2 I(V, V).$$

Therefore we obtain

$$\liminf_{s \to 0} \frac{f(\xi_v(\varepsilon s)) + f(\xi_v(-\varepsilon s)) - 2f(x)}{\varepsilon^2 s^2}$$
$$\leq \liminf_{s \to 0} \frac{L(\sigma_s)^2 + L(\sigma_{-s})^2 - 2L(\eta)^2}{2\varepsilon^2 s^2} = \frac{1}{2\varepsilon^2} \left. \frac{\partial^2 L(\sigma_s)}{\partial s^2} \right|_{s=0}$$
$$\leq -\varepsilon^{-1} g_{\dot\eta(0)}(v, v)/d(x,y) + \mathcal{T}_{\dot\eta(0)}(v)/2d(x,y) + I(V, V)/2.$$

Letting ε go to zero completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.2 Volume distortion coefficients

The next lemma plays a key role in the proof of Proposition 5.3. Recall (1.2) for the definition of volume distortion coefficients $\mathfrak{v}_t^<$ and $\mathfrak{v}_t^>$. For a linear operator $Q : T_x M \longrightarrow T_y M$, we define $\mathbf{D}[Q] := \mathsf{m}_y(Q(A))/\mathsf{m}_x(A)$, where m_x and m_y stand for Lebesgue measures on $T_x M$ and $T_y M$ induced from m, and $A \subset T_x M$ is an arbitrary nonempty, bounded open set.

Lemma 3.2 *Fix distinct points $x, y \in M$ with $y \notin \operatorname{Cut}(x)$ and let $\eta : [0, 1] \longrightarrow M$ be the unique minimal geodesic from x to y. For $t \in (0, 1]$, we define $\bar{f}_t(z) := -d(z, \eta(t))^2/2$. Then we have, for any $t \in (0, 1)$,*

$$\mathfrak{v}_t^<(x, y) = \mathbf{D}\big[d(\exp_x)_{\nabla \bar{f}_t(x)} \circ [d(\exp_x)_{\nabla \bar{f}_1(x)}]^{-1}\big],$$

$$\mathfrak{v}_t^>(x, y) = (1 - t)^{-n}\mathbf{D}\Big[d(\exp_x \circ \mathfrak{L}_x)_{d(\bar{f}_t)_x} \circ \big[d\big(d(t\bar{f}_1)\big)_x - d(d\bar{f}_t)_x\big]\Big].$$

In the first equation, we identify $T_{\nabla \bar{f}_t(x)}(T_x M)$ and $T_{\nabla \bar{f}_1(x)}(T_x M)$. In the second, the vertical part of $T_{d(\bar{f}_t)_x}(T^* M)$ is identified with $T_{d(\bar{f}_t)_x}(T_x^* M)$.

Proof. For small $r > 0$, we define $\tau_t^< := \exp_x \circ(t \exp_x^{-1}) : B^+(y, r) \longrightarrow M$. Then we have

$$\mathfrak{v}_t^<(x, y) = \lim_{r \to 0} \frac{\mathsf{m}(\tau_t^<(B^+(y, r)))}{\mathsf{m}(B^+(y, tr))} = t^{-n}\mathbf{D}[d(\tau_t^<)_y] = \mathbf{D}\big[d(\exp_x)_{\nabla \bar{f}_t(x)} \circ [d(\exp_x)_{\nabla \bar{f}_1(x)}]^{-1}\big].$$

We similarly define $\tau_t^> := \exp_y^{\bar{F}} \circ((1 - t)(\exp_y^{\bar{F}})^{-1}) : B^-(x, r) \longrightarrow M$, where $\exp^{\bar{F}}$ denotes the exponential map with respect to the reverse Finsler structure $\bar{F}$ of F. We can rewrite this as $\tau_t^>(z) = \exp_z \circ \mathfrak{L}_z(t \cdot d(\bar{f}_1)_z)$, and hence

$$\mathfrak{v}_t^>(x, y) = \lim_{r \to 0} \frac{\mathsf{m}(\tau_t^>(B^-(x, r)))}{\mathsf{m}(B^-(x, (1 - t)r))} = (1 - t)^{-n}\mathbf{D}[d(\tau_t^>)_x]$$

$$= (1 - t)^{-n}\mathbf{D}\big[d(\exp \circ \mathfrak{L})_{d(t\bar{f}_1)_x} \circ d\big(d(t\bar{f}_1)\big)_x\big].$$

In order to synchronize with the assertion of the lemma, we remark that $\exp_z \circ \mathfrak{L}_z(d(\bar{f}_t)_z) = \eta(t)$ for all $z \in B^-(x, r)$, so that $d(\exp \circ \mathfrak{L})_{d(\bar{f}_t)_x} \circ d(d\bar{f}_t)_x = 0$. Since $d(\bar{f}_t)_x = d(t\bar{f}_1)_x$, we obtain

$$\mathfrak{v}_t^>(x, y) = (1 - t)^{-n}\mathbf{D}\Big[d(\exp \circ \mathfrak{L})_{d(\bar{f}_t)_x} \circ \big[d\big(d(t\bar{f}_1)\big)_x - d(d\bar{f}_t)_x\big]\Big].$$

Furthermore, as the image of $d(d(t\bar{f}_1))_x - d(d\bar{f}_t)_x : T_x M \longrightarrow T_{d(\bar{f}_t)_x}(T^* M)$ is contained in the vertical part (see (2.1)), we can replace $d(\exp \circ \mathfrak{L})_{d(\bar{f}_t)_x}$ with $d(\exp_x \circ \mathfrak{L}_x)_{d(\bar{f}_t)_x}$. $\qquad \square$

We remark that the function $h(z) := t\bar{f}_1(z) - \bar{f}_t(z) = d(z, \eta(t))^2/2 - td(z, y)^2/2$ satisfies

$$\left(\frac{\partial^2 h}{\partial x^i \partial x^j}(x)\right) \geq 0 \tag{3.1}$$

in any local coordinate system around x. In fact, we have

$$2h(z) \geq d\big(z, \eta(t)\big)^2 - t\big\{d\big(z, \eta(t)\big) + d\big(\eta(t), y\big)\big\}^2$$

$$= (1 - t)d\big(z, \eta(t)\big)^2 - td\big(\eta(t), y\big)^2 - 2\sqrt{1 - t}\frac{t}{\sqrt{1 - t}}d\big(z, \eta(t)\big)d\big(\eta(t), y\big)$$

$$\geq -\left(\frac{t^2}{1 - t} + t\right)d\big(\eta(t), y\big)^2 = -(1 - t)td(x, y)^2 = 2h(x).$$

4 Optimal transport via c-concave functions

In this section, as the first step toward Theorem 1.1, we study how optimal transport between two measures is described using a c-concave function. Our discussion closely follows McCann's [Mc] concerning Riemannian manifolds. See [Br] for the case of Euclidean spaces and [Be] for Alexandrov spaces. We also refer to [AGS], [RR], [Vi1] and [Vi2] for background information and further developments. In this and the following sections, c always stands for the *quadratic cost function* $c(x,y) := d(x,y)^2/2$ on a Finsler manifold (M,F), and $\bar{c}(y,x) := c(x,y)$ is its reverse.

4.1 c-concave functions

Let $X, Y \subset M$ be two compact sets. Given an arbitrary function $\phi : X \longrightarrow \mathbb{R} \cup \{-\infty\}$, we define its *$c$-transform* $\phi^c : Y \longrightarrow \mathbb{R} \cup \{-\infty\}$ relative to (X,Y) by

$$\phi^c(y) := \inf_{x \in X} \{c(x,y) - \phi(x)\}.$$

We similarly define the $\bar{c}$-transform of a function $\psi : Y \longrightarrow \mathbb{R} \cup \{-\infty\}$ relative to (Y,X) by

$$\psi^{\bar{c}}(x) := \inf_{y \in Y} \{\bar{c}(y,x) - \psi(y)\} = \inf_{y \in Y} \{c(x,y) - \psi(y)\}$$

for $x \in X$. Be careful of the order of x and y.

Definition 4.1 (*c-concave functions*) Let $X, Y \subset M$ be two compact sets. Then a function $\phi : X \longrightarrow \mathbb{R} \cup \{-\infty\}$ is said to be *c-concave* relative to (X,Y) if it is not identically $-\infty$ and if there is a function $\psi : Y \longrightarrow \mathbb{R} \cup \{-\infty\}$ whose $\bar{c}$-transform $\psi^{\bar{c}}$ relative to (Y,X) coincides with ϕ.

We also say that $\psi : Y \longrightarrow \mathbb{R} \cup \{-\infty\}$ is *$\bar{c}$-concave* relative to (Y,X) if it is not identically $-\infty$ and if there is a function $\phi : X \longrightarrow \mathbb{R} \cup \{-\infty\}$ for which $\phi^c = \psi$ holds. We will always use (X,Y) for ϕ and (Y,X) for ψ, so that they are sometimes omitted. We summarize basic facts of the c-transform and c-concave functions in the next lemma.

Lemma 4.2 *Take two compact sets $X, Y \subset M$ and a function $\phi : X \longrightarrow \mathbb{R} \cup \{-\infty\}$. Then the following properties hold*:

(i) *We have $\phi \leq \phi^{c\bar{c}}$ and $\phi^c = \phi^{c\bar{c}c}$.*

(ii) *Assume that ϕ is not identically $-\infty$. Then ϕ is c-concave if and only if $\phi = \phi^{c\bar{c}}$.*

(iii) *If ϕ is c-concave, then it is Lipschitz continuous and its Lipschitz constant is bounded above by a constant depending only on X and Y.*

In particular, given a c-concave function $\phi : X \longrightarrow \mathbb{R}$ and $x \in X$, we can choose a point $y \in Y$ which attains $\inf_{y \in Y}\{c(x,y) - \phi^c(y)\}$ $(= \phi^{c\bar{c}}(x) = \phi(x))$.

Definition 4.3 (*c*-superdifferentials) Let $X, Y \subset M$ be compact sets and $\phi : X \longrightarrow \mathbb{R}$ be a *c*-concave function relative to (X, Y). Then the *c-superdifferential* of ϕ at a point $x \in X$ is the nonempty set

$$\partial^c \phi(x) := \{ y \in Y \mid \phi(x) = c(x, y) - \phi^c(y) \}.$$

For a $\bar{c}$-concave function $\psi : Y \longrightarrow \mathbb{R}$ relative to (Y, X) and $y \in Y$, we similarly define

$$\partial^{\bar{c}} \psi(y) := \{ x \in X \mid \psi(y) = c(x, y) - \psi^{\bar{c}}(x) \}.$$

We record two straightforward properties for later convenience.

Lemma 4.4 *Take two compact sets $X, Y \subset M$ and a c-concave function $\phi : X \longrightarrow \mathbb{R}$. Then the following properties hold*:

(i) *It holds that $y \in \partial^c \phi(x)$ if and only if $x \in \partial^{\bar{c}} \phi^c(y)$.*

(ii) *We have $y \in \partial^c \phi(\partial^{\bar{c}} \phi^c(y))$.*

Proof. (i) Note that both $y \in \partial^c \phi(x)$ and $x \in \partial^{\bar{c}} \phi^c(y)$ are equivalent to $\phi(x) + \phi^c(y) = c(x, y)$. (ii) Take $x \in \partial^{\bar{c}} \phi^c(y)$ and observe that (i) implies $y \in \partial^c \phi(x) \subset \partial^c \phi(\partial^{\bar{c}} \phi^c(y))$. $\square$

Since a *c*-concave function ϕ is Lipschitz continuous, it is differentiable almost everywhere. At a differentiable point, we find a nice description of $\partial^c \phi(x)$.

Lemma 4.5 *Take two compact sets $X, Y \subset M$ and a c-concave function $\phi : X \longrightarrow \mathbb{R}$. If ϕ is differentiable at $x \in X$, then we have $\partial^c \phi(x) = \{ \exp_x(\nabla(-\phi)(x)) \}$. Moreover, the curve $\eta(t) := \exp_x(t \nabla(-\phi)(x))$, $t \in [0, 1]$, is a unique minimal geodesic from x to $\exp_x(\nabla(-\phi)(x))$.*

Proof. Fix arbitrary $y \in \partial^c \phi(x)$ and define $f(z) := c(z, y) = d(z, y)^2 / 2$. For any $v \in T_x M$, the definition of $\partial^c \phi(x)$ yields that

$$f(\exp_x v) \geq \phi^c(y) + \phi(\exp_x v) = f(x) - \phi(x) + \phi(\exp_x v) = f(x) + d\phi_x(v) + o(F(v)).$$

(Thus $d\phi_x \in \partial_* f(x)$ in the sense of Definition 4.6 below.) Take a minimal geodesic $\eta : [0, d(x, y)] \longrightarrow M$ from x to y. Given $\varepsilon > 0$, put $y_\varepsilon = \eta(d(x, y) - \varepsilon)$ and note that $\eta|_{[0, d(x,y)-\varepsilon]}$ does not cross the cut locus of x. Then we observe from the first variation formula (2.7) that, as in the proof of Lemma 3.1,

$$
\begin{aligned}
f(\exp_x v) - f(x) &\leq \{ d(\exp_x v, y_\varepsilon) + \varepsilon \}^2 / 2 - \{ d(x, y_\varepsilon) + \varepsilon \}^2 / 2 \\
&= -\{ d(x, y_\varepsilon) + \varepsilon \} g_{\dot\eta(0)}(v, \dot\eta(0)) + o(F(v)) \\
&= -d(x, y) \mathfrak{L}_x^{-1}(\dot\eta(0))(v) + o(F(v)).
\end{aligned}
$$

This implies $d\phi_x(v) \leq -d(x, y) \mathfrak{L}_x^{-1}(\dot\eta(0))(v)$ for all $v \in T_x M$, and hence $\nabla(-\phi)(x) = d(x, y) \cdot \dot\eta(0)$. Therefore we obtain $y = \eta(d(x, y)) = \exp_x(\nabla(-\phi)(x))$ and η is the unique minimal geodesic from x to y. $\square$

The above lemma will be sharpened up in Proposition 5.1 using the almost everywhere second order differentiability of c-concave functions. Such differentiability was established in [Oh4] along with a generalized Alexandrov-Bangert theorem. To state it, we recall terminologies in non-smooth analysis.

Definition 4.6 (Subdifferentials) Take a function $f : M \longrightarrow \mathbb{R}$ and a point $x \in M$. Then a co-vector $\alpha \in T_x^* M$ is called a *subgradient* of f at x if we have

$$f(\exp_x v) \geq f(x) + \alpha(v) + o\big(F(v)\big)$$

for $v \in T_x M$. The set of all subgradients at x is called the *subdifferential* of f at x and denoted by $\partial_* f(x) \subset T_x^* M$.

Any c-concave function admits a (not necessarily unique) subgradient everywhere. In Euclidean spaces and Riemannian manifolds, the dual of α in Definition 4.6 is usually called the subgradient, and the corresponding subdifferential $\partial f(x)$ is a subset of $T_x M$. Also in Finsler manifolds, it is possible to define a subgradient as an element of $T_x M$ through the Legendre transform. It actually coincides with the gradient vector of f if f is differentiable at x, namely $\partial_* f(x) = \{df_x\}$ and $\partial f(x) = \{\nabla f(x)\}$.

Definition 4.7 (Second order differentials) Take a function $f : M \longrightarrow \mathbb{R}$ and let $\Phi : U \longrightarrow \mathbb{R}^n$ be a local coordinate system on an open set $U \subset M$ with $\Phi(x) = 0$. Then f is said to be *second order differentiable* at x if f is differentiable at x and if there is a linear map $H : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ such that

$$\sup_{\alpha(z) \in \partial_* f(z)} \left\| \big[d(\Phi^{-1})_u\big]^* \big(\alpha(z)\big) - \big[d(\Phi^{-1})_0\big]^* (df_x) - Hu \right\| = o(\|u\|)$$

for $u = \Phi(z) \in \mathbb{R}^n$, where we identify $T_u^* \mathbb{R}^n$, $T_0^* \mathbb{R}^n$ and $\mathbb{R}^n$ in the left-hand side.

At the level of the local coordinate system $\Phi = (x^i)_{i=1}^n$, we can write

$$H = \left(\frac{\partial^2 f}{\partial x^i \partial x^j}(x) \right).$$

Hence we are able to define $d(df)_x : T_x M \longrightarrow T_{df_x}(T^* M)$ as in (2.1).

Theorem 4.8 ([Oh4, Theorem 7.4]) *Take a compact set $Y \subset M$ and an open set $U \subset M$ whose closure $X := \overline{U}$ is compact. Then any c-concave function $\phi : X \longrightarrow \mathbb{R}$ relative to (X, Y) is second order differentiable* m-*a.e. on U.*

In particular, $d\phi : U \longrightarrow T^* M$ is continuous on the domain consisting of second order differentiable points of ϕ.

4.2 The Brenier-McCann solution to the Monge-Kantorovich problem

We define $\mathcal{P}(M)$ as the set of Borel probability measures on M and $\mathcal{P}_c(M) \subset \mathcal{P}(M)$ as the subset consisting of compactly supported measures. Given $\mu, \nu \in \mathcal{P}_c(M)$, the Monge

problem is to find (and characterize) a map $\mathcal{F} : M \longrightarrow M$ which attains the infimum of $\int_M c(x, \mathcal{F}(x)) \, d\mu(x)$ among all maps pushing μ forward to ν (we write it $\mathcal{F}_{\#}\mu = \nu$). Kantorovich shed new light on the problem using the set

$$\Pi(\mu, \nu) := \{\pi \in \mathcal{P}(M \times M) \,|\, (p_1)_{\#}\pi = \mu, \ (p_2)_{\#}\pi = \nu\},$$

where p_1 and p_2 are projections to the first and second entries. Each $\pi \in \Pi(\mu, \nu)$ is called a *coupling* of (μ, ν) and the Monge-Kantorovich problem is to find an *optimal coupling* attaining $\inf_{\pi \in \Pi(\mu,\nu)} \int_{M \times M} c \, d\pi$.

Brenier [Br] showed that a unique solution to the Monge-Kantorovich problem (and simultaneously to the Monge problem) in a Euclidean space is reprensented as the gradient of a convex function, and McCann [Mc] successfully extended it to Riemannian manifolds using c-concave functions. McCann's technique is applicable to Finsler manifolds, and it is more or less equivalent to proving the *Kantorovich duality*:

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{M \times M} c(x, y) \, d\pi(x, y) = \sup_{(\phi,\psi) \in \mathrm{Lip}^c(X,Y)} \left\{ \int_M \phi \, d\mu + \int_M \psi \, d\nu \right\}. \qquad (4.1)$$

Here $X \supset \mathrm{supp}\,\mu$, $Y \supset \mathrm{supp}\,\nu$ and $\mathrm{Lip}^c(X, Y)$ stands for the set of pairs of Lipschitz continuous functions $(\phi, \psi) \in \mathrm{Lip}(X) \times \mathrm{Lip}(Y)$ satisfying $\phi(x) + \psi(y) \le c(x, y)$ for all $(x, y) \in X \times Y$. Thus the inequality $\ge$ is clear in (4.1). The duality itself is known to hold true in much more general situations (see [AGS, §6.1], [Vi2, Chapter 5]).

We first observe that a c-concave function naturally appears as a maximizer of the right-hand side of (4.1).

Lemma 4.9 *Let $\mu, \nu \in \mathcal{P}_c(M)$ and take compact sets $X \supset \mathrm{supp}\,\mu$ and $Y \supset \mathrm{supp}\,\nu$. Then there exists a c-concave function $\phi : X \longrightarrow \mathbb{R}$ such that $(\phi, \phi^c) \in \mathrm{Lip}^c(X, Y)$ and it attains the supremum in the right-hand side of (4.1). Moreover, if μ is absolutely continuous, then the vector field $\nabla(-\phi)$ is unique among such maximizers.*

Proof. Given $(\phi, \psi) \in \mathrm{Lip}^c(X, Y)$, we first observe that $(\phi^{c\bar{c}}, \phi^c) \in \mathrm{Lip}^c(X, Y)$, $\phi^c \ge \psi$ and $\phi^{c\bar{c}} \ge \phi$.

Take a maximizing sequence $\{(\phi_i, \psi_i)\}_{i \in \mathbb{N}} \subset \mathrm{Lip}^c(X, Y)$ and a point $x_0 \in X$. Then the sequence $\{(\hat{\phi}_i, \hat{\psi}_i)\} := \{(\phi_i^{c\bar{c}} - \phi_i^{c\bar{c}}(x_0), \phi_i^c + \phi_i^{c\bar{c}}(x_0))\} \subset \mathrm{Lip}^c(X, Y)$ is also maximizing and $\hat{\phi}_i$ is c-concave with $(\hat{\phi}_i)^c = \hat{\psi}_i$. Since $\hat{\phi}_i(x_0) = 0$ and $\hat{\phi}_i$ is Lipschitz continuous with a uniformly bounded Lipschitz constant (Lemma 4.2(iii)), the Ascoli-Arzelà theorem provides a subsequence of $\{\hat{\phi}_i\}$ which converges uniformly to a Lipschitz function $\phi : X \longrightarrow \mathbb{R}$. We also find that the corresponding subsequence of $\{\hat{\psi}_i\}$ converges to ϕ^c and $\phi^{c\bar{c}} = \lim_{i \to \infty}(\hat{\phi}_i)^{c\bar{c}} = \phi$. Thus ϕ is c-concave and (ϕ, ϕ^c) attains the supremum in (4.1) by construction.

In order to prove the uniqueness, we take two maximizing pairs $(\phi_1, \psi_1), (\phi_2, \psi_2) \in \mathrm{Lip}^c(X, Y)$. Then we deduce that ϕ_i is c-concave and $\psi_i = \phi_i^c$ by the above discussion. Put $\phi = (\phi_1 + \phi_2)/2$ and note that $\phi^c \ge (\phi_1^c + \phi_2^c)/2$ on Y. As $(\phi, \phi^c) \in \mathrm{Lip}^c(X, Y)$, this implies that $\phi^c = (\phi_1^c + \phi_2^c)/2$ and ϕ is c-concave. Moreover, if $y \in \partial^c\phi(x)$, then we observe $y \in \partial^c\phi_1(x) \cap \partial^c\phi_2(x)$. Thus it follows from Lemma 4.5 together with the absolute continuity of μ that $\nabla(-\phi_1)(x) = \nabla(-\phi_2)(x) = \nabla(-\phi)(x)$ for μ-a.e. $x \in X$. $\qquad \square$

Now we give a precise description of a unique solution to the Monge and the Monge-Kantorovich problems.

Theorem 4.10 *Given $\mu, \nu \in \mathcal{P}_c(M)$ such that μ is absolutely continuous with respect to m, take a compact set $Y \supset \operatorname{supp}\nu$ and an open set $U \supset \operatorname{supp}\mu$ whose closure $X := \overline{U}$ is compact. Then there exists a c-concave function $\phi : X \longrightarrow \mathbb{R}$ relative to (X, Y) such that $\pi := (\operatorname{Id}_M \times \mathcal{F})_{\#}\mu$ is a unique optimal coupling of (μ, ν), where we define $\mathcal{F}(x) := \exp_x(\nabla(-\phi)(x))$. Moreover, $\mathcal{F}$ is a unique optimal transport map from μ to ν.*

Proof. By Lemma 4.9, there exists a c-concave function $\phi : X \longrightarrow \mathbb{R}$ for which (ϕ, ϕ^c) attains the supremum in (4.1). We define $\mathcal{F}(x) := \exp_x(\nabla(-\phi)(x))$ for $x \in X$ at where ϕ is differentiable, and remark that the domain has μ-full measure since μ is absolutely continuous. Recall also that $\mathcal{F}$ is continuous on an m-full measure subset of U (Theorem 4.8), so that $\mathcal{F}$ is measurable on X.

We shall show that $\mathcal{F}_{\#}\mu = \nu$. Take a continuous function $h \in C(M)$ and put $\psi_\varepsilon = \phi^c + \varepsilon h$ for $\varepsilon \in \mathbb{R}$ close to 0. Given $x \in X$, we find a point $y_\varepsilon \in Y$ such that $c(x, y_\varepsilon) - \psi_\varepsilon(y_\varepsilon) = (\psi_\varepsilon)^{\bar{c}}(x)$. It follows from Lemma 4.5 that, if ϕ is differentiable at x, then y_ε tends to $y_0 = \mathcal{F}(x)$ as ε goes to zero. Combining this with

$$\phi(x) - \varepsilon h(y_\varepsilon) \leq c(x, y_\varepsilon) - \phi^c(y_\varepsilon) - \varepsilon h(y_\varepsilon) = (\psi_\varepsilon)^{\bar{c}}(x)$$
$$\leq c(x, \mathcal{F}(x)) - \psi_\varepsilon(\mathcal{F}(x)) = \phi(x) - \varepsilon h(\mathcal{F}(x)),$$

we find $(\psi_\varepsilon)^{\bar{c}}(x) = \phi(x) - \varepsilon h(\mathcal{F}(x)) + o(|\varepsilon|)$ as well as $|o(|\varepsilon|)| \leq 2\varepsilon \|h|_Y\|_\infty$. We put $J(\varepsilon) = \int_M (\psi_\varepsilon)^{\bar{c}} \, d\mu + \int_M \psi_\varepsilon \, d\nu$ and deduce from the maximality of (ϕ, ϕ^c) that

$$0 = \lim_{\varepsilon \to 0} \frac{J(\varepsilon) - J(0)}{\varepsilon} = -\int_M h \, d[\mathcal{F}_{\#}\mu] + \int_M h \, d\nu,$$

and hence $\mathcal{F}_{\#}\mu = \nu$.

Define $\pi_\phi := (\operatorname{Id}_M \times \mathcal{F})_{\#}\mu \in \Pi(\mu, \nu)$ and observe that $c(x, y) = \phi(x) + \phi^c(y)$ holds π_ϕ-a.e. by Lemma 4.5. Thus $\int_{M \times M} c \, d\pi_\phi = \int_M \phi \, d\mu + \int_M \phi^c \, d\nu$, so that π_ϕ is an optimal coupling. (Now we finished the proof of the Kantorovich duality (4.1).) Conversely, thanks to (4.1), any optimal coupling $\pi \in \Pi(\mu, \nu)$ must satisfy $c(x, y) = \phi(x) + \phi^c(y)$ π-a.e. It means that $\pi(\bigcup_{x \in X}(x, \mathcal{F}(x))) = 1$, therefore we conclude that $\pi = (\operatorname{Id}_M \times \mathcal{F})_{\#}\mu = \pi_\phi$. $\square$

We show one corollary to the above proof for later convenience.

Corollary 4.11 *Take a compact set $Y \subset M$ and an open set $U \subset M$ whose closure $X := \overline{U}$ is compact. Then, for any c-concave function $\phi : X \longrightarrow \mathbb{R}$ and any absolutely continuous measure $\mu \in \mathcal{P}_c(M)$ with $\operatorname{supp}\mu \subset U$, the map $\mathcal{F}(x) := \exp_x(\nabla(-\phi)(x))$ is the unique optimal transport map from μ to $\mathcal{F}_{\#}\mu$.*

Proof. Put $\nu = \mathcal{F}_{\#}\mu$ and take the unique optimal coupling $\pi \in \Pi(\mu, \nu)$ given by Theorem 4.10. Then, as in the proof of Theorem 4.10, we deduce from Lemma 4.5 that

$$\int_{M \times M} c \, d\pi \geq \int_M \phi \, d\mu + \int_M \phi^c \, d\nu = \int_{M \times M} c \, d[(\operatorname{Id}_M \times \mathcal{F})_{\#}\mu] \geq \int_{M \times M} c \, d\pi.$$

Thus we have $c(x, y) = \phi(x) + \phi^c(y)$ π-a.e., and hence $\pi = (\operatorname{Id}_M \times \mathcal{F})_{\#}\mu$. $\square$

4.3 Wasserstein spaces

We briefly explain a geometric conception behind Theorem 4.10. For $\mu, \nu \in \mathcal{P}_c(M)$, the L_2-*Wasserstein distance* is defined by

$$d_2^W(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{M \times M} d(x, y)^2 d\pi(x, y) \right)^{1/2}. \tag{4.2}$$

Note that $d_2^W(\mu, \nu)^2/2$ is the left-hand side of (4.1). If μ is absolutely continuous, then Theorem 4.10 provides the unique optimal coupling $(\mathrm{Id}_M \times \mathcal{F})_{\#}\mu$. Moreover, the curve $\mu_t := (\mathcal{F}_t)_{\#}\mu$ with $\mathcal{F}_t(x) := \exp_x(\nabla(-t\phi)(x))$, $t \in [0,1]$, turns out a unique minimal geodesic in $(\mathcal{P}_c(M), d_2^W)$ from μ to ν. Indeed, it is clear by construction that

$$\int_{M \times M} d(x, y)^2 \, d[(\mathcal{F}_s \times \mathcal{F}_t)_{\#}\mu](x, y) = (t - s)^2 \int_M d\big(x, \mathcal{F}(x)\big)^2 \, d\mu(x) = (t - s)^2 d_2^W(\mu, \nu)^2$$

for any $0 \le s < t \le 1$, so that $(\mu_t)_{t \in [0,1]}$ is minimal. Uniqueness can be seen as follows. Denote by $\Gamma(M) \subset \mathrm{Lip}([0,1], M)$ the space of minimal geodesics $\gamma : [0,1] \longrightarrow M$ equipped with the uniform topology, and define $e_t : \Gamma(M) \longrightarrow M$ by $e_t(\gamma) := \gamma(t)$ for $t \in [0,1]$. Given a minimal geodesic $\alpha : [0,1] \longrightarrow \mathcal{P}_c(M)$ from μ to ν, there is a Borel probability measure Ξ on $\Gamma(M)$ such that $(e_t)_{\#}\Xi = \alpha(t)$ for all $t \in [0,1]$ and that $(e_0 \times e_1)_{\#}\Xi$ is an optimal coulping of (μ, ν) (see [LV1, Proposition 2.10]). Hence we have $(e_0 \times e_1)_{\#}\Xi = (\mathrm{Id}_M \times \mathcal{F})_{\#}\mu$. This together with Lemma 4.5 shows that $\Xi = [\nabla(-\phi)]_{\#}\mu$, where we abbreviated $[\nabla(-\phi)](x) := (\mathcal{F}_t(x))_{t \in [0,1]} \in \Gamma(M)$. Therefore we obtain $\alpha(t) = \mu_t$.

5 The Jacobian equation

As the second step toward Theorem 1.1, we analyze the optimal transport map $\mathcal{F}$ given in Theorem 4.10. The present and the next sections are based on and generalize a Riemannian discussion in [CMS1].

5.1 Differentiating optimal transport maps

By virtue of Theorem 4.8, a c-concave function is almost everywhere second order differentiable. At such a nice point x, we can differentiate the map $\mathcal{F}$ and its differential is represented by using the second order differentials of ϕ and of the distance function from $\mathcal{F}(x)$.

Proposition 5.1 *Take a compact set $Y \subset M$ and an open set $U \subset M$ whose closure $X := \overline{U}$ is compact, and let $\phi : X \longrightarrow \mathbb{R}$ be a c-concave function. Define $\mathcal{F}(z) := \exp_z(\nabla(-\phi)(z))$ and fix $x \in U$ at where ϕ is second order differentiable and $d\phi_x \ne 0$. Then we have the following:*

(i) *The point $y := \mathcal{F}(x)$ is not a cut point of x.*

(ii) *The function $h(z) := c(z, y) - \phi(z)$ satisfies $dh_x = 0$ and*

$$\left(\frac{\partial^2 h}{\partial x^i \partial x^j}(x) \right) \geq 0$$

in any local coordinate system $(x^i)_{i=1}^n$ around x.

(iii) *Define $\bar{f}(z) := -c(z, y)$ and*

$$d\mathcal{F}_x := d(\exp_x \circ \mathfrak{L}_x)_{d(-\phi)_x} \circ \left[d\big(d(-\phi)\big)_x - d(d\bar{f})_x \right] : T_x M \longrightarrow T_y M,$$

where we identify the vertical part of $T_{d(-\phi)_x}(T^*M)$ and $T_{d(-\phi)_x}(T^*_x M)$. Then we have, for $v \in T_x M$,

$$\sup \big\{ F\big(u - d\mathcal{F}_x(v)\big) \mid \exp_y u \in \partial^c \phi(\exp_x v),\, F(u) = d(y, \exp_y u) \big\} = o\big(F(v)\big).$$

Proof. (i) First of all, it follows from Lemma 4.5 that $\partial^c \phi(x) = \{y\}$. Hence we have, for any unit vector $v \in T_x M$ and small $t > 0$,

$$c(x, y) - \phi(x) = \phi^c(y) \leq c\big(\xi_v(\pm t), y\big) - \phi\big(\xi_v(\pm t)\big), \tag{5.1}$$

where $\xi_v : (-\varepsilon, \varepsilon) \longrightarrow M$ is the geodesic with $\dot{\xi}_v(0) = v$. Thus we put $f(z) = c(z, y)$ and observe

$$\frac{\phi(\xi_v(t)) + \phi(\xi_v(-t)) - 2\phi(x)}{t^2} \leq \frac{f(\xi_v(t)) + f(\xi_v(-t)) - 2f(x)}{t^2}.$$

As ϕ is second order differentiable at x, this implies

$$\frac{\partial^2 (\phi \circ \xi_v)}{\partial t^2}(0) \leq \liminf_{t \to 0} \frac{f(\xi_v(t)) + f(\xi_v(-t)) - 2f(x)}{t^2}.$$

The left-hand side admits a uniform lower bound in v, and hence y is not a cut point of x by Lemma 3.1.

(ii) This is clear because $\bar{f} = -c(\cdot, y)$ is C^∞ at x (recall $y \neq x$), $\nabla \bar{f}(x) = \nabla(-\phi)(x)$ and because (5.1) shows that h takes a minimum at x.

(iii) We first remark that $dh_x = 0$ in (ii) ensures that $d(-\phi)_x = d\bar{f}_x$, and hence the difference $d\big(d(-\phi)\big)_x - d(d\bar{f})_x$ makes sense. Fix a unit vector $v \in T_x M$, put $x_t = \exp_x tv$ for small $t \geq 0$ and take $u_t \in T_y M$ such that $y_t := \exp_y u_t \in \partial^c \phi(x_t)$ as well as $d(y, y_t) = F(u_t)$. We define $\bar{f}_t(z) := -c(z, y_t)$ and see that $d(\bar{f}_t)_{x_t} \in \partial_*(-\phi)(x_t)$ since $y_t \in \partial^c \phi(x_t)$ implies that, as in (5.1),

$$-\phi(\exp_{x_t} w) \geq -\phi(x_t) - \bar{f}_t(x_t) + \bar{f}_t(\exp_{x_t} w) = -\phi(x_t) + d(\bar{f}_t)_{x_t}(w) + o\big(F(w)\big)$$

for $w \in T_{x_t} M$. Hence differentiating $y_t = \exp \circ \mathfrak{L}(d(\bar{f}_t)_{x_t})$ at $t = 0$ gives (recall Definition 4.7)

$$\left. \frac{\partial y_t}{\partial t} \right|_{t=0} = d(\exp \circ \mathfrak{L})_{d(-\phi)_x} \circ d\big(d(-\phi)\big)_x(v).$$

Moreover, it follows from $\exp \circ \mathfrak{L}(d\bar{f}_{x_t}) \equiv y$ that $d(\exp \circ \mathfrak{L})_{d\bar{f}_x} \circ [d(d\bar{f})_x](v) = 0$. Therefore we have

$$\left. \frac{\partial y_t}{\partial t} \right|_{t=0} = d(\exp \circ \mathfrak{L})_{d(-\phi)_x} \circ \left[d\big(d(-\phi)\big)_x - d(d\bar{f})_x \right](v) = d\mathcal{F}_x(v).$$

19

In the last equality, as in Lemma 3.2, the image of $d(d(-\phi))_x - d(d\bar{f})_x$ contains only vertical terms (see (5.2) below), thus we could regard it as living in $T_{d(-\phi)_x}(T_x^* M)$ and replace $d(\exp \circ \mathfrak{L})_{d(-\phi)_x}$ with $d(\exp_x \circ \mathfrak{L}_x)_{d(-\phi)_x}$. We consequently obtain $\exp_y u_t = y_t = \exp_y(d\mathcal{F}_x(tv) + o(t))$ and the error term $o(t)$ is uniform in v because of the second order differentiability of ϕ at x. This completes the proof. $\qquad\square$

We excluded the case of $d\phi_x = 0$ from the above proposition for the fact that the function $c(\cdot, x)$ is only C^1 at x unless F is Riemannian. At the level of a local coordinate system (see (2.1)), we can write

$$d\big(d(-\phi)\big)_x - d(d\bar{f})_x = \sum_{i,j=1}^{n} \frac{\partial^2 h}{\partial x^i \partial x^j}(x) \frac{\partial}{\partial v^i}\bigg|_{d(-\phi)_x} dx^j|_x. \qquad (5.2)$$

We can check that (5.2) is indeed coordinate-free taking $dh_x = 0$ into account.

5.2   The Jacobian equation and concavity

We are next concerned with quantitative behavior of $d\mathcal{F}$. We shall see that the Jacobian equation holds, that is to say, the ratio of density functions of μ and $\mathcal{F}_{\#}\mu$ coincides with $\mathbf{D}[d\mathcal{F}]$ (defined as in the paragraph preceding Lemma 3.2).

Theorem 5.2 *Given two absolutely continuous measures $\mu = f\mathsf{m}, \nu = g\mathsf{m} \in \mathcal{P}_c(M)$, take open sets $U \supset \operatorname{supp}\mu, W \supset \operatorname{supp}\nu$ whose closures $X := \overline{U}, Y := \overline{W}$ are compact. Denote by $\phi : X \longrightarrow \mathbb{R}$ a c-concave function as in Theorem 4.10 and define $\mathcal{F}(x) := \exp_x(\nabla(-\phi)(x))$. Then $\mathcal{F}$ is injective on a μ-full measure set and we have the following for μ-a.e. $x \in U \setminus (d\phi)^{-1}(0)$.*

(i) *The function $h(z) := c(z, \mathcal{F}(x)) - \phi(z)$ satisfies*

$$\left(\frac{\partial^2 h}{\partial x^i \partial x^j}(x)\right) > 0$$

in any local coordinate system $(x^i)_{i=1}^n$ around x. In other words, $\mathbf{D}[d\mathcal{F}_x] > 0$ holds for $d\mathcal{F}_x : T_x M \longrightarrow T_{\mathcal{F}(x)} M$ defined as in Proposition 5.1.

(ii) *We have $\lim_{r \to 0} \mathsf{m}(\partial^c \phi(B^+(x,r)))/\mathsf{m}(B^+(x,r)) = \mathbf{D}[d\mathcal{F}_x]$.*

(iii) *It holds that $f(x) = g(\mathcal{F}(x))\mathbf{D}[d\mathcal{F}_x]$.*

Proof. Define $\Omega_X \subset U$ as the set of points $x \in X$ such that ϕ is second order differentiable at x as well as x is a Lebesgue point of f. Similarly, we define $\Omega_Y \subset W$ as the set of Lebesgue points $y \in Y$ of g at where ϕ^c is second order differentiable. We find $\mu(\Omega_X) = \nu(\Omega_Y) = 1$ by Theorem 4.8 together with the absolute continuity of μ and ν. Thus we define $\Omega := \Omega_X \cap \mathcal{F}^{-1}(\Omega_Y)$ and observe $\mu(\Omega) = 1$. It follows from Theorem 4.10 that $\mathcal{F}^*(y) := \exp_y^{\bar{F}}(\nabla^{\bar{F}}(-\phi^c)(y))$ is the unique optimal transport map pushing ν forward to μ in the sense that $(\mathcal{F}^* \times \operatorname{Id}_M)_{\#}\nu$ is the unique optimal coupling of (μ, ν). Here $\exp^{\bar{F}}$ is the exponential map with respect to the reverse $\bar{F}$ of F. Moreover, $\mathcal{F}^* \circ \mathcal{F} = \operatorname{Id}_X$ holds

on Ω because $\{\mathcal{F}^* \circ \mathcal{F}(x)\} = \partial^{\bar{c}}\phi^c(\mathcal{F}(x)) \ni x$ (see Lemmas 4.4(i), 4.5). In particular, $\mathcal{F}$ is injective on Ω.

We will show (i)–(iii) for $x \in \Omega$ with $d\phi_x \neq 0$. We observe from $\mathcal{F}^* \circ \mathcal{F} = \mathrm{Id}_X$ that $d\phi_x \neq 0$ is equivalent to $d\phi^c_{\mathcal{F}(x)} \neq 0$. We put $y = \mathcal{F}(x)$ in the following.

(i) This follows from $d\mathcal{F}^*_y \circ d\mathcal{F}_x = \mathrm{Id}_{T_xM}$ combined with Proposition 5.1(ii) (see also (5.2)). Given a unit vector $v \in T_xM$ and small $t > 0$, we choose $u_t \in T_yM$ so as to satisfy $\exp_y u_t \in \partial^c\phi(\exp_x tv)$ and $d(y, \exp_y u_t) = F(u_t)$. Note that this implies $\exp_x tv \in \partial^{\bar{c}}\phi^c(\exp_y u_t)$ by Lemma 4.4(i). Then applying Proposition 5.1(iii) twice shows that

$$ tv = d\mathcal{F}^*_y(u_t) + o(t) = d\mathcal{F}^*_y \circ d\mathcal{F}_x(tv) + o(t). $$

Thus $v = d\mathcal{F}^*_y \circ d\mathcal{F}_x(v)$.

(ii) We deduce from Proposition 5.1(iii) for $v \in B^+_{T_xM}(0, r)$ along with the invertibility of $d\mathcal{F}_x$ that

$$ \partial^c\phi\big(B^+(x, r)\big) \subset \exp_y\left[d\mathcal{F}_x\big(B^+_{T_xM}(0, r)\big) + B^+_{T_yM}\big(0, o(r)\big)\right] $$
$$ \subset \exp_y\left[d\mathcal{F}_x\Big(B^+_{T_xM}\big(0, r + o(r)\big)\Big)\right]. $$

Moreover, the same discussion for $v \in d\mathcal{F}_x(B^+_{T_xM}(0, r))$ and ϕ^c yields that

$$ \partial^{\bar{c}}\phi^c\Big(\exp_y\left[d\mathcal{F}_x\big(B^+_{T_xM}(0, r)\big)\right]\Big) \subset \exp_x\left[B^+_{T_xM}(0, r) + B^+_{T_xM}\big(0, o(r)\big)\right] $$
$$ \subset B^+\big(x, r + o(r)\big). $$

Combining this with Lemma 4.4(ii) shows

$$ \partial^c\phi\Big(B^+\big(x, r + o(r)\big)\Big) \subset \exp_y\left[d\mathcal{F}_x\big(B^+_{T_xM}(0, r)\big)\right] \subset \partial^c\phi\Big(B^+\big(x, r + o(r)\big)\Big). \qquad (5.3) $$

Therefore we conclude that

$$ \lim_{r \to 0} \frac{\mathsf{m}(\partial^c\phi(B^+(x, r)))}{\mathsf{m}(B^+(x, r))} = \lim_{r \to 0} \frac{\mathsf{m}_y(d\mathcal{F}_x(B^+_{T_xM}(0, r)))}{\mathsf{m}_x(B^+_{T_xM}(0, r))} = \mathbf{D}[d\mathcal{F}_x]. $$

(iii) As x is a Lebesgue point of f, we find $f(x) = \lim_{r \to 0} \mu(B^+(x, r))/\mathsf{m}(B^+(x, r))$. Moreover, (5.3) and Lemma 4.5 together with the injectivity of $\mathcal{F}$ imply

$$ g(y) = \lim_{r \to 0} \frac{\nu(\partial^c\phi(B^+(x, r)))}{\mathsf{m}(\partial^c\phi(B^+(x, r)))} = \lim_{r \to 0} \frac{\nu(\mathcal{F}(B^+(x, r)))}{\mathsf{m}(\partial^c\phi(B^+(x, r)))} = \lim_{r \to 0} \frac{\mu(B^+(x, r))}{\mathsf{m}(\partial^c\phi(B^+(x, r)))}. $$

We consequently obtain

$$ f(x) = \lim_{r \to 0} \left\{ \frac{\mu(B^+(x, r))}{\mathsf{m}(\partial^c\phi(B^+(x, r)))} \frac{\mathsf{m}(\partial^c\phi(B^+(x, r)))}{\mathsf{m}(B^+(x, r))} \right\} = g(y)\mathbf{D}[d\mathcal{F}_x]. $$

$\square$

We next prove the concavity of $\mathbf{D}[d\mathcal{F}]^{1/n}$ along the optimal transport from μ to ν. This is a key estimate in the proof of Theorem 1.1. Although it was not emphasized in [CMS1], the use of the classical (linear) Brunn-Minkowski inequality in the proof seems interesting. Compare this with generalized (nonlinear) Brunn-Minkowski inequalities in Corollaries 9.3, 9.4(ii) and 9.5(ii).

Proposition 5.3 *Take a compact set $Y \subset M$ and an open set $U \subset M$ whose closure $X := \overline{U}$ is compact, and let $\phi : X \longrightarrow \mathbb{R}$ be a c-concave function. Fix $x \in U$ at where ϕ is second order differentiable with $d\phi_x \neq 0$ and, for $t \in (0,1]$, define $y_t := \exp_x(\nabla(-t\phi))(x)$, $\bar{f}_t(z) := -c(z, y_t)$ and $\mathbf{J}_t(x) := \mathbf{D}[d(\mathcal{F}_t)_x]$, where*

$$d(\mathcal{F}_t)_x := d(\exp_x \circ \mathfrak{L}_x)_{d(-t\phi)_x} \circ \left[d\big(d(-t\phi)\big)_x - d(d\bar{f}_t)_x \right] : T_x M \longrightarrow T_{y_t} M.$$

Then we have, for any $t \in (0,1)$,

$$\mathbf{J}_t(x)^{1/n} \geq (1-t)\mathfrak{v}_t^{>}(x, y_1)^{1/n} + t\mathfrak{v}_t^{<}(x, y_1)^{1/n}\mathbf{J}_1(x)^{1/n}.$$

Proof. We start with

$$d\big(d(-t\phi)\big)_x - d(d\bar{f}_t)_x = \left\{ d\big(d(-t\phi)\big)_x - d\big(d(t\bar{f}_1)\big)_x \right\} + \left\{ d\big(d(t\bar{f}_1)\big)_x - d(d\bar{f}_t)_x \right\}.$$

Note that $d(\bar{f}_t)_x = d(-t\phi)_x = d(t\bar{f}_1)_x$. We introduce the operator $\tau_s : T^*M \longrightarrow T^*M$ as $\tau_s(\alpha) = s\alpha$ for $s > 0$, and observe

$$
\begin{aligned}
&d(\exp_x \circ \mathfrak{L}_x)_{d(-t\phi)_x} \circ \left[d\big(d(-t\phi)\big)_x - d\big(d(t\bar{f}_1)\big)_x \right] \\
&= d(\exp_x \circ \mathfrak{L}_x)_{d(-t\phi)_x} \circ d(\tau_t)_{d(-\phi)_x} \circ \left[d\big(d(-\phi)\big)_x - d(d\bar{f}_1)_x \right] \\
&= d(\exp_x \circ \mathfrak{L}_x)_{d(-t\phi)_x} \circ d(\tau_t)_{d(-\phi)_x} \circ \left[d(\exp_x \circ \mathfrak{L}_x)_{d(-\phi)_x} \right]^{-1} \circ d(\mathcal{F}_1)_x \\
&= d(\exp_x)_{\nabla(-t\phi)(x)} \circ d(\mathfrak{L}_x \circ \tau_t \circ \mathfrak{L}_x^{-1})_{\nabla(-\phi)(x)} \circ \left[d(\exp_x)_{\nabla(-\phi)(x)} \right]^{-1} \circ d(\mathcal{F}_1)_x \\
&= t \cdot d(\exp_x)_{\nabla(-t\phi)(x)} \circ \left[d(\exp_x)_{\nabla(-\phi)(x)} \right]^{-1} \circ d(\mathcal{F}_1)_x.
\end{aligned}
$$

In the last line, we identified $T_{\nabla(-t\phi)(x)}(T_xM)$ and $T_{\nabla(-\phi)(x)}(T_xM)$. Therefore the classical Brunn-Minkowski inequality in $T_{y_t}M$ and Lemma 3.2 show that

$$
\begin{aligned}
\mathbf{J}_t(x)^{1/n} &= \mathbf{D}[d(\mathcal{F}_t)_x]^{1/n} \\
&\geq (1-t)\mathbf{D}\left[(1-t)^{-1} d(\exp_x \circ \mathfrak{L}_x)_{d(-t\phi)_x} \circ \left[d\big(d(t\bar{f}_1)\big)_x - d(d\bar{f}_t)_x \right] \right]^{1/n} \\
&\quad + t\mathbf{D}\left[d(\exp_x)_{\nabla(-t\phi)(x)} \circ \left[d(\exp_x)_{\nabla(-\phi)(x)} \right]^{-1} \circ d(\mathcal{F}_1)_x \right]^{1/n} \\
&= (1-t)\mathfrak{v}_t^{>}(x, y_1)^{1/n} + t\mathfrak{v}_t^{<}(x, y_1)^{1/n}\mathbf{J}_1(x)^{1/n}.
\end{aligned}
$$

$\square$

6 A Finsler Borell-Brascamp-Lieb inequality

This section contains the proof of Theorem 1.1.

6.1 Absolute continuity of intermediate measures

We first discuss two qualitative properties of optimal transport maps. See [Vi2, Chapter 8] for more comprehensive treatment.

Lemma 6.1 *Take a compact set $Y \subset M$ and an open set $U \subset M$ whose closure $X := \overline{U}$ is compact, and let $\phi : X \longrightarrow \mathbb{R}$ be a c-concave function. Define $\mathcal{F}_t(x) := \exp_x(\nabla(-t\phi)(x))$ for $t \in [0,1]$ and fix $t \in (0,1)$. Then $\mathcal{F}_t$ is injective on a subset of U of m-full measure.*

Proof. Let $\Omega \subset U$ be the set of second order differentiable points of ϕ. Then Theorem 4.8 ensures that $\mathsf{m}(\Omega) = \mathsf{m}(U)$ and $\mathcal{F}_s$ is continuous on Ω for every $s \in [0,1]$. We will prove that $\mathcal{F}_t$ is injective on Ω. Assume that there are two distinct points $x, x' \in \Omega$ such that $\mathcal{F}_t(x) = \mathcal{F}_t(x')$, and put $z = \mathcal{F}_t(x) = \mathcal{F}_t(x')$, $y = \mathcal{F}_1(x)$ and $y' = \mathcal{F}_1(x')$. Then we have

$$
\begin{aligned}
d(x,y')^2 + d(x',y)^2 &\leq \{d(x,z) + d(z,y')\}^2 + \{d(x',z) + d(z,y)\}^2 \\
&= \{d(x,z) + d(z,y)\}^2 + \{d(x',z) + d(z,y')\}^2 \\
&\quad + 2\{d(x,z)d(z,y') + d(x',z)d(z,y) - d(x,z)d(z,y) - d(x',z)d(z,y')\} \\
&= d(x,y)^2 + d(x',y')^2 - 2(1-t)t\{d(x,y) - d(x',y')\}^2 \\
&\leq d(x,y)^2 + d(x',y')^2.
\end{aligned}
$$

If strict inequality

$$
d(x,y')^2 + d(x',y)^2 < d(x,y)^2 + d(x',y')^2 \tag{6.1}
$$

holds, then the continuity of $\mathcal{F}_1$ shows that $\mathcal{F}_1$ is not an optimal transport map from $\mu := \mathsf{m}(B^+(x,r) \cup B^+(x',r))^{-1}\mathsf{m}|_{B^+(x,r) \cup B^+(x',r)}$ to $\nu := (\mathcal{F}_1)_{\#}\mu$ for sufficiently small $r > 0$. In fact, (6.1) implies $y \neq y'$, and it is more efficient to transport $\mu|_{B^+(x,r)}$ to $\nu|_{\mathcal{F}_1(B^+(x'r))}$ as well as $\mu|_{B^+(x',r)}$ to $\nu|_{\mathcal{F}_1(B^+(x,r))}$. This contradicts Corollary 4.11 and hence we have

$$
d(x,y')^2 + d(x',y)^2 = d(x,y)^2 + d(x',y')^2. \tag{6.2}
$$

Therefore equality holds in the above inequality, thus we see that $d(x,y) = d(x',y')$. As z is the common $t : (1-t)$-fraction point, (6.2) shows that $x = x'$ and $y = y'$ which contradict our choice $x \neq x'$. $\qquad \square$

Lemma 6.2 *Given two absolutely continuous measures $\mu = f\mathsf{m}, \nu = g\mathsf{m} \in \mathcal{P}_c(M)$, take open sets $U \supset \operatorname{supp}\mu, W \supset \operatorname{supp}\nu$ whose closures $X := \overline{U}, Y := \overline{W}$ are compact. Denote by $\phi : X \longrightarrow \mathbb{R}$ a c-concave function as in Theorem 4.10 and define $\mathcal{F}_t(x) := \exp_x(\nabla(-t\phi)(x))$ for $t \in [0,1]$. Then, for any $t \in (0,1)$, the measure $\mu_t := (\mathcal{F}_t)_{\#}\mu$ is absolutely continuous with respect to m.*

Proof. We first recall that $\mathcal{F}_t$ is injective on a μ-full measure set by Lemma 6.1 along with the absolute continuity of μ. Denote by $U' \subset U$ the set of points x with $d\phi_x = 0$. Note that $\mu_t|_{U'} = (\mathcal{F}_t)_{\#}(\mu|_{U'}) = \mu|_{U'}$ is clearly absolutely continuous. Combining Theorems 4.8 and 5.2(i), we find $\Omega \subset U \setminus U'$ such that $\mu(\Omega) = \mu(U \setminus U')$, $\mathbf{D}[d(\mathcal{F}_1)_x] > 0$ for $x \in \Omega$ and that $\mathcal{F}_t$ is continuous on Ω. Here $d(\mathcal{F}_t)_x : T_x M \longrightarrow T_{\mathcal{F}_t(x)}M$ is defined by, as in Proposition 5.3,

$$
d(\mathcal{F}_t)_x := d(\exp_x \circ \mathcal{L}_x)_{d(-t\phi)_x} \circ \left[d\big(d(-t\phi)\big)_x - d(d\bar{f}_t)_x \right],
$$

where $\bar{f}_s(z) := -c(z, \mathcal{F}_s(x))$ for $s \in (0,1]$. Given $x \in \Omega$, note that

$$
d\big(d(-t\phi)\big)_x - d(d\bar{f}_t)_x = \left\{ d\big(d(-t\phi)\big)_x - d\big(d(t\bar{f}_1)\big)_x \right\} + \left\{ d\big(d(t\bar{f}_1)\big)_x - d(d\bar{f}_t)_x \right\}.
$$

Thus we have $\mathbf{D}[d(\mathcal{F}_t)_x] > 0$ by $\mathbf{D}[d(\mathcal{F}_1)_x] > 0$ and (3.1).

23

For each $x \in \Omega$, the invertibility of $d(\mathcal{F}_t)_x$ and the Taylor expansion in Proposition 5.1(iii) imply that $d(\mathcal{F}_t(x), \mathcal{F}_t(z)) \geq a_x d(x, z)$ holds for some $a_x, r_x > 0$ and all $z \in B^+(x, r_x) \cap \Omega$. Since $\mathcal{F}_t$ is continuous, a_x and r_x can be taken lower semi-continuously in Ω. Therefore, for each $x \in \Omega$, there exist $a'_x, r'_x > 0$ such that $(\mathcal{F}_t)^{-1}$ is $(a'_x)^{-1}$-Lipschitz continuous on $\mathcal{F}_t(B^+(x, r'_x) \cap \Omega)$. Combining this with the absolute continuity of μ shows that $\mu_t|_{\mathcal{F}_t(B^+(x,r'_x)\cap\Omega)} = (\mathcal{F}_t)_\#(\mu|_{B^+(x,r'_x)\cap\Omega})$ is absolutely continuous. Thus we obtain the absolute continuity of $\mu_t|_{\mathcal{F}_t(\Omega)}$ and complete the proof. $\qquad\square$

Now we are ready to finish proving Theorem 1.1.

6.2   Proof of Theorem 1.1

We first discuss the case where both A and B are bounded. We put $\mu = f|_A\mathsf{m}$, $\nu = g|_B\mathsf{m}$ and take open sets $U \supset A$, $W \supset B$ whose closures $X := \overline{U}$, $Y := \overline{W}$ are compact. By virtue of Theorem 4.10, we find a c-concave function $\phi : X \longrightarrow \mathbb{R}$ such that the map $\mathcal{F}_s(x) := \exp_x(\nabla(-s\phi)(x))$, $s \in [0, 1]$, gives the unique optimal transport from μ to ν. Lemma 6.2 ensures that $(\mathcal{F}_t)_\#\mu$ is absolutely continuous, so that we can write $(\mathcal{F}_t)_\#\mu = \rho_t\mathsf{m}$. Define $U' \subset U$ as the set of points x at where $d\phi_x$ is defined and $d\phi_x = 0$. We further choose $\Omega_t, \Omega_1 \subset U \setminus U'$ such that $\mu(\Omega_t) = \mu(\Omega_1) = \mu(U \setminus U')$, f is positive on Ω_1 and that all conclusions in Theorem 5.2 hold true for $\mathcal{F} = \mathcal{F}_t, \mathcal{F}_1$, respectively. We finally set $\Omega = \Omega_t \cap \Omega_1 \cap A \cap \mathcal{F}_1^{-1}(B)$. Recall that both $\mathcal{F}_t$ and $\mathcal{F}_1$ are injective by Lemma 6.1 and Theorem 5.2.

We deduce from Theorem 5.2(iii) that

$$f(x) = \rho_t\big(\mathcal{F}_t(x)\big)\mathbf{D}[d(\mathcal{F}_t)_x] = g\big(\mathcal{F}_1(x)\big)\mathbf{D}[d(\mathcal{F}_1)_x] > 0$$

for $x \in \Omega$. Thus it follows from Proposition 5.3 that

$$\frac{1}{\rho_t(\mathcal{F}_t(x))^{1/n}} \geq (1-t)\left(\frac{\mathfrak{v}_t^>(x, \mathcal{F}_1(x))}{f(x)}\right)^{1/n} + t\left(\frac{\mathfrak{v}_t^<(x, \mathcal{F}_1(x))}{g(\mathcal{F}_1(x))}\right)^{1/n}.$$

Combining this with the hypothesis on h, we obtain $h(\mathcal{F}_t(x)) \geq \rho_t(\mathcal{F}_t(x))$ for $x \in \Omega$. Therefore we have

$$\int_{\mathcal{F}_t(\Omega)} h \, d\mathsf{m} \geq \int_{\mathcal{F}_t(\Omega)} \rho_t \, d\mathsf{m} = \int_\Omega d\mu = \mu(\Omega).$$

For $x \in \Omega' := U' \cap A \cap B$, we observe $\mathcal{F}_1(x) = x$ and $f|_{\Omega'}\mathsf{m} = (\mathcal{F}_1)_\#(f|_{\Omega'}\mathsf{m}) = g|_{\Omega'}\mathsf{m}$. Hence we find, for a Lebesgue point $x \in \Omega'$ of both $f|_{\Omega'}$ and $g|_{\Omega'}$ such that $f(x) > 0$,

$$\frac{1}{h(x)^{1/n}} \leq \frac{1-t}{f(x)^{1/n}} + \frac{t}{g(x)^{1/n}} = \frac{1}{f(x)^{1/n}}.$$

We used the fact that $\mathfrak{v}_t^>(x, x) = \mathfrak{v}_t^<(x, x) = 1$. This shows $\int_{\Omega'} h \, d\mathsf{m} \geq \int_{\Omega'} d\mu = \mu(\Omega')$. As $\mathsf{m}(\mathcal{F}_t(\Omega) \cap \Omega') = 0$ by Lemma 6.1, we conclude that

$$\int_M h \, d\mathsf{m} \geq \int_{\mathcal{F}_t(\Omega)} h \, d\mathsf{m} + \int_{\Omega'} h \, d\mathsf{m} \geq \mu(\Omega \cup \Omega') = 1.$$

If either A or B is unbounded, then we choose bounded subsets $A_\varepsilon \subset A$ and $B_\varepsilon \subset B$ such that $\int_{A_\varepsilon} f \, d\mathsf{m} = \int_{B_\varepsilon} g \, d\mathsf{m} = 1 - \varepsilon$ for $\varepsilon > 0$. Applying the above discussion to $f/(1-\varepsilon)$ and $g/(1-\varepsilon)$ yields $\int_M h/(1-\varepsilon) \, d\mathsf{m} \geq 1$. As $\varepsilon > 0$ was arbitrary, we eventually obtain $\int_M h \, d\mathsf{m} \geq 1$ and complete the proof. $\qquad\square$

7 Volume comparison theorems

We proceed to the second main part, volume comparison theorems under Ricci curvature bounds. We first define two families of functions used to describe the theorems. For $K \in \mathbb{R}$, $N \in (1, \infty)$ and $r \in (0, \infty)$ ($r \in (0, \pi\sqrt{(N-1)/K})$ if $K > 0$), we define

$$
\mathbf{s}_{K,N}(r) := \begin{cases} \sqrt{(N-1)/K} \sin(r\sqrt{K/(N-1)}) & \text{if } K > 0, \\ r & \text{if } K = 0, \\ \sqrt{-(N-1)/K} \sinh(r\sqrt{-K/(N-1)}) & \text{if } K < 0. \end{cases}
$$

In addition, for $t \in (0, 1)$, we define

$$
\beta_{K,N}^t(r) := \left(\frac{\mathbf{s}_{K,N}(tr)}{t\mathbf{s}_{K,N}(r)} \right)^{N-1}, \qquad \beta_{K,\infty}^t(r) := e^{K(1-t^2)r^2/6}. \tag{7.1}
$$

Here the infinite dimensional case ($N = \infty$) is derived as the limit and will be necessary in the next section. By computation, we also observe (see [vRS, §2])

$$
\log\left[\beta_{K,N}^{1-t}(r)^{1-t} \cdot \beta_{K,N}^t(r)^t \right] \geq \frac{K}{2}(1-t)tr^2 \tag{7.2}
$$

for all $N \in (1, \infty]$ (equality holds if $N = \infty$).

Recall (1.2) for the definition of volume distortion coefficients $\mathfrak{v}_t^<$ and $\mathfrak{v}_t^>$, and (1.4) and (1.5) for the definition of $\mathcal{V}$, $\partial_v \mathcal{V}$ and $\partial_v^2 \mathcal{V}$. In this context, Shen's volume comparison theorem is stated as follows. It generalizes the Bishop and the Bishop-Gromov comparison theorems in Riemannian geometry.

Theorem 7.1 ([Sh1, Theorem 1.1], [Sh2, Theorem 16.1.1]) *Assume that $n \geq 2$ and there are constants $K \in \mathbb{R}$ and $H \geq 0$ such that*

$$
\mathrm{Ric}(v) \geq K, \qquad \partial_v \mathcal{V} \geq -H
$$

holds for all unit vectors $v \in TM$. Then we have, for any $x \in M$, $y \in M \setminus \mathrm{Cut}(x)$ and $t \in (0, 1)$,

$$
\mathfrak{v}_t^<(x, y) \geq \frac{e^{tHd(x,y)}}{e^{Hd(x,y)}} \beta_{K,n}^t\big(d(x,y)\big), \qquad \mathfrak{v}_t^>(x, y) \geq \frac{e^{(1-t)Hd(x,y)}}{e^{Hd(x,y)}} \beta_{K,n}^{1-t}\big(d(x,y)\big).
$$

In particular, for any $0 < r < R$ ($\leq \pi\sqrt{(n-1)/K}$ if $K > 0$), it holds that

$$
\max\left\{ \frac{\mathsf{m}(B^+(x, R))}{\mathsf{m}(B^+(x, r))}, \frac{\mathsf{m}(B^-(x, R))}{\mathsf{m}(B^-(x, r))} \right\} \leq \frac{\int_0^R e^{tH} \mathbf{s}_{K,n}(t)^{n-1} \, dt}{\int_0^r e^{tH} \mathbf{s}_{K,n}(t)^{n-1} \, dt}.
$$

We remark that, by virtue of a generalized Bonnet-Myers theorem due to Auslander [Au], we know that $\mathrm{Ric}(v) \geq K > 0$ implies $\mathrm{diam}\, M \leq \pi\sqrt{(n-1)/K}$ (see also [BCS, Theorem 7.7.1]).

Now we present a different volume comparison theorem inspired by the theory of weighted Riemannian manifolds. See [BE], [Qi] and [Lo] for related analytic, geometric and topological work on weighted Riemannian manifolds. Our discussion is based on a strategy used in [LV1] for proving N-Ricci curvature bounds of weighted Riemannian manifolds (see also [Qi]).

Lemma 7.2 *Fix a unit speed geodesic $\eta : [0, l) \longrightarrow M$ which does not cross the cut locus $\mathrm{Cut}(\eta(0))$. Take an orthonormal basis $\{e_i\}_{i=1}^n$ of $T_x M$ with respect to $g_{\dot{\eta}(0)}$ such that $e_n = \dot{\eta}(0)$, and define $E_i(r) := d(\exp_{\eta(0)})_{r\dot{\eta}(0)}(re_i) \in T_{\eta(r)}M$ for $r \in [0, l)$. Then we have, for any $r \in (0, l)$,*

$$\frac{h''(r)}{h(r)} \leq -\frac{\mathrm{Ric}(\dot{\eta}(r))}{n-1}, \quad \text{where } h(r) := \left\{ \det \left[\left(g_{\dot{\eta}(r)}(E_i, E_j) \right)_{i,j=1}^{n-1} \right] \right\}^{1/2(n-1)}.$$

Proof. We first remark that every E_i is a Jacobi field by construction. Throughout the proof, we denote $D_{\dot{\eta}}^{\dot{\eta}} E_i(r)$ by $E_i'(r)$ and put $g_r = g_{\dot{\eta}(r)}$ for brevity. Define an $(n-1) \times (n-1)$ matrix $A(r) = (a_{ij}(r))$ by $a_{ij}(r) := g_r(E_i, E_j)$. Thanks to the Finsler version of the Gauss lemma (cf. [BCS, Lemma 6.6.1]), we have $g_r(E_i, \dot{\eta}) \equiv 0$ which with (2.5) implies $g_r(E_i', \dot{\eta}) \equiv 0$ for all $i = 1, 2, \ldots, n-1$. Since $\eta(r) \notin \mathrm{Cut}(\eta(0))$, for each $r \in (0, l)$, we find an $(n-1) \times (n-1)$ matrix $B(r) = (b_{ij}(r))$ such that $E_i'(r) = \sum_{j=1}^{n-1} b_{ij}(r) E_j(r)$. Note that $A' = BA + AB^t$.

It follows from (2.3) that

$$\frac{d}{dr}\left[g_r(E_i', E_j) - g_r(E_i, E_j') \right] = -g_r\left(R^{\dot{\eta}}(E_i, \dot{\eta})\dot{\eta}, E_j \right) + g_r\left(E_i, R^{\dot{\eta}}(E_j, \dot{\eta})\dot{\eta} \right) = 0,$$

so that $g_r(E_i', E_j) = g_r(E_i, E_j')$. Thus we see that $BA = AB^t$ as well as $A' = 2BA$. Note also that

$$a_{ij}''(r) = -g_r\left(R^{\dot{\eta}}(E_i, \dot{\eta})\dot{\eta}, E_j \right) - g_r\left(E_i, R^{\dot{\eta}}(E_j, \dot{\eta})\dot{\eta} \right) + 2g_r(E_i', E_j'),$$

and hence $A'' = -2\,\mathrm{Ric}_{\dot{\eta}(r)} + 2BAB^t = -2\,\mathrm{Ric}_{\dot{\eta}(r)} + 2B^2 A$, where we set

$$(\mathrm{Ric}_{\dot{\eta}(r)})_{ij} := g_r\left(R^{\dot{\eta}}(E_i, \dot{\eta})\dot{\eta}, E_j \right).$$

We observe $h(r) = \{\det A(r)\}^{1/2(n-1)}$ and

$$(n-1)h' = \frac{1}{2}(\det A)'(\det A)^{1/2(n-1)-1} = \frac{h}{2}\,\mathrm{trace}(A'A^{-1}),$$

$$(n-1)h'' = h'\,\mathrm{trace}\,B + \frac{h}{2}\,\mathrm{trace}\left(A''A^{-1} - (A'A^{-1})^2 \right)$$

$$= \frac{h}{n-1}(\mathrm{trace}\,B)^2 - h\,\mathrm{trace}(\mathrm{Ric}_{\dot{\eta}(r)}\,A^{-1}) - h\,\mathrm{trace}(B^2).$$

Take another $(n-1) \times (n-1)$ matrix $C(r) = (c_{ij}(r))$ for which $\{\sum_{j=1}^{n-1} c_{ij}(r)E_j(r)\}_{i=1}^{n-1} \cup \{\dot{\eta}(r)\}$ forms an orthonormal basis of $T_{\eta(r)}M$ with respect to g_r. Then we find $CAC^t = I_{n-1}$ and hence $\mathrm{trace}(\mathrm{Ric}_{\dot{\eta}(r)}\,A^{-1}) = \mathrm{trace}(C\,\mathrm{Ric}_{\dot{\eta}(r)}\,C^t) = \mathrm{Ric}(\dot{\eta}(r))$. Moreover, plugging $A^{-1} = C^t C$ into $BA = AB^t$ shows that $CBC^{-1} = (CBC^{-1})^t$. This implies that

$$(\mathrm{trace}\,B)^2 = \{\mathrm{trace}(CBC^{-1})\}^2 \leq (n-1)\,\mathrm{trace}\left((CBC^{-1})^2 \right) = (n-1)\,\mathrm{trace}(B^2).$$

Therefore we obtain $(n-1)h''(r) \leq -h(r)\,\mathrm{Ric}(\dot{\eta}(r))$ as required. $\qquad\square$

Theorem 7.3 (A volume comparison theorem) *Assume that $n \geq 2$ and there are constants $K \in \mathbb{R}$ and $N \in (n, \infty)$ such that*

$$\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} - \frac{1}{N-n}(\partial_v \mathcal{V})^2 \geq K$$

holds for all unit vectors $v \in TM$. Then we have, for any $x \in M$, $y \in M \setminus \mathrm{Cut}(x)$ and $t \in (0,1)$ (provided that $d(x,y) < \pi\sqrt{(N-1)/K}$ if $K > 0$),

$$\mathfrak{v}_t^<(x,y) \geq t^{N-n}\beta_{K,N}^t\big(d(x,y)\big), \qquad \mathfrak{v}_t^>(x,y) \geq (1-t)^{N-n}\beta_{K,N}^{1-t}\big(d(x,y)\big).$$

In particular, we have $\mathrm{diam}\, M \leq \pi\sqrt{(N-1)/K}$ if $K > 0$ and, for any $0 < r < R$ ($\leq \pi\sqrt{(N-1)/K}$ if $K > 0$), it holds that

$$\max\left\{\frac{\mathsf{m}(B^+(x,R))}{\mathsf{m}(B^+(x,r))}, \frac{\mathsf{m}(B^-(x,R))}{\mathsf{m}(B^-(x,r))}\right\} \leq \frac{\int_0^R \mathsf{s}_{K,N}(t)^{N-1}\, dt}{\int_0^r \mathsf{s}_{K,N}(t)^{N-1}\, dt}.$$

Proof. As proof is common, we will treat only $\mathfrak{v}_t^<$ and B^+. Fix a unit vector $v \in T_xM$ and let $\eta : [0,l] \longrightarrow M$ be the geodesic with $\dot\eta(0) = v$ which does not cross $\mathrm{Cut}(x)$. Define $f(r) := r^n \mathbf{D}[d(\exp_x)_{rv}]$ for $r \in [0,l]$ by identifying $T_{rv}(T_xM)$ with T_xM. Then the first assertion of the theorem is equivalent to that the function $r \longmapsto f(r)/r\mathsf{s}_{K,N}(r)^{N-1}$ is nonincreasing. For simplicity, we set $\mathcal{V}(r) = \mathcal{V}(\dot\eta(r))$ and $g_r = g_{\dot\eta(r)}$ in this proof.

We first remark that

$$f(r) = r^n \frac{\mathbf{D}[d(\exp_x)_{rv}]}{\mathbf{D}_g[d(\exp_x)_{rv}]}\mathbf{D}_g[d(\exp_x)_{rv}] = e^{\mathcal{V}(0)-\mathcal{V}(r)}r^n\mathbf{D}_g[d(\exp_x)_{rv}],$$

where $\mathbf{D}_g$ is taken with respect to g_0 and g_r. Take an orthonormal basis $\{e_i\}_{i=1}^n$ of T_xM with respect to g_v such that $e_n = v$. We extend these to vector fields $E_i(r) := d(\exp_x)_{rv}(re_i)$ along η. Note that $E_n(r) = r\dot\eta(r)$. We put

$$h(r) := \left(\frac{f(r)}{re^{\mathcal{V}(0)}}\right)^{1/(N-1)} = e^{-\mathcal{V}(r)/(N-1)}\left\{\det\left[\big(g_r(E_i, E_j)\big)_{i,j=1}^{n-1}\right]\right\}^{1/2(N-1)}.$$

We used the Gauss lemma in the second equality as in the proof of Lemma 7.2. We divide h into two parts, that is,

$$h_1(r) := e^{-\mathcal{V}(r)/(N-n)}, \qquad h_2(r) := \left\{\det\left[\big(g_r(E_i, E_j)\big)_{i,j=1}^{n-1}\right]\right\}^{1/2(n-1)}.$$

Note that $h = h_1^{(N-n)/(N-1)}h_2^{(n-1)/(N-1)}$. We calculate, using Lemma 7.2,

$$(N-1)h^{-1}h'' = (N-n)h_1^{-1}h_1'' + (n-1)h_2^{-1}h_2'' - \frac{(N-n)(n-1)}{N-1}(h_1^{-1}h_1' - h_2^{-1}h_2')^2$$

$$\leq -\partial_{\dot\eta(r)}^2\mathcal{V} + \frac{1}{N-n}(\partial_{\dot\eta(r)}\mathcal{V})^2 - \mathrm{Ric}\big(\dot\eta(r)\big).$$

Thus the hypothesis implies $h^{-1}h'' \leq -K/(N-1)$. Comparing this with $\mathsf{s}_{K,N}^{-1}\mathsf{s}_{K,N}'' = -K/(N-1)$ shows that $h(r)/\mathsf{s}_{K,N}(r) = \{f(r)/re^{\mathcal{V}(0)}\mathsf{s}_{K,N}(r)^{N-1}\}^{1/(N-1)}$ is nonincreasing.

The second assertion follows from the first because any minimal geodesic of length $\pi\sqrt{(N-1)/K}$ must contain a conjugate point. The third assertion follows by integration in a standard manner (see, e.g., [Ch, Lemma III.4.1]). $\square$

On a Riemannian manifold (M, g), a weighted measure is usually represented as the form $e^{-V} \mathrm{vol}_g$, where V is a function on M and the reference measure vol_g is the Riemannian volume element. In contrast, our reference measure in the above proof is the measure induced from $g_{\dot{\eta}(r)}$ and it depends on v. If (M, F) is Berwald type, then the Busemann-Hausdorff measure m_{BH} satisfies $\partial_v \mathcal{V} \equiv 0$ ([Sh1, Propositions 2.6, 2.7]), and hence m_{BH} can be thought of as a reference measure. However, it seems impossible to find such a measure in general. In this sense, every measure is weighted, and it is natural to consider an arbitrary measure m in the first place.

Theorem 7.3 is regarded as the *measure contraction property* in the sense of [Oh1], [Oh2] and [St3], and it is a corollary to the curvature-dimension condition (Theorem 1.2(i)), though we have presented above a simpler, direct proof.

8 The curvature-dimension condition

Inspired by [CMS1] and [OV] as well, von Renesse and Sturm [vRS] showed that the K-convexity of the relative entropy is equivalent to $\mathrm{Ric} \geq K$ for Riemannian manifolds. Sturm [St1], [St2], [St3] and Lott and Villani [LV1], [LV2] independently developed this theory introducing the curvature-dimension condition and the N-Ricci curvature bound, respectively. They proved that these (essentially equivalent) conditions characterize Ricci curvature bounds of Riemannian manifolds, as we will extend to Finsler manifolds. Furthermore, general (symmetric) metric measure spaces satisfying these conditions are known to behave like spaces with lower Ricci curvature bounds (see Corollaries 8.4, 8.5 for instance). We refer to [Oh1], [Oh2], [vR] for related work based on the measure contraction property, [Ol1], [Ol2] for another approach using the contraction of Markov chains, and to [Vi2, Part III] for further developments.

8.1 Definition

In order to define the N-Ricci curvature bound, we introduce important classes of functions (see [LV1], [LV2] for details). For $N \in [1, \infty)$, denote by $\mathcal{DC}_N$ the set of continuous convex functions $U : [0, \infty) \longrightarrow \mathbb{R}$ such that $U(0) = 0$ and that the function $\varphi(s) := s^N U(s^{-N})$ is convex on $(0, \infty)$. We remark that the convexity of U and $U(0) = 0$ imply that, for $0 < s < t$,

$$\varphi(t) \leq t^N \left\{ \frac{s^{-N} - t^{-N}}{s^{-N}} U(0) + \frac{t^{-N}}{s^{-N}} U(s^{-N}) \right\} = \varphi(s).$$

Thus φ is nonincreasing. We similarly define $\mathcal{DC}_\infty$ as the set of continuous convex functions $U : [0, \infty) \longrightarrow \mathbb{R}$ such that $U(0) = 0$ and that $\varphi(s) := e^s U(e^{-s})$ is convex on $\mathbb{R}$. This φ is also nonincreasing. For an absolutely continuous measure $\mu = \rho \mathsf{m} \in \mathcal{P}(M)$, we define $U_\mathsf{m}(\mu) := \int_M U(\rho) \, d\mathsf{m}$. We remark that $\mathcal{DC}_{N'} \subset \mathcal{DC}_N$ if $N < N'$.

The most important element of $\mathcal{DC}_N$ is $U(r) = Nr(1 - r^{-1/N})$ which derives the *Rényi entropy* $U_\mathsf{m}(\rho \mathsf{m}) = N - N \int_M \rho^{1-1/N} \, d\mathsf{m}$. Letting N go to infinity provides $U(r) = r \log r \in \mathcal{DC}_\infty$ as well as the *relative entropy* $\mathrm{Ent}_\mathsf{m}(\rho \mathsf{m}) := U_\mathsf{m}(\rho \mathsf{m}) = \int_M \rho \log \rho \, d\mathsf{m}$.

By Theorem 4.10, for any two absolutely continuous measures $\mu_0, \mu_1 \in \mathcal{P}_c(M)$, there is a unique minimal geodesic $(\mu_t)_{t\in[0,1]}$ from μ_0 to μ_1 in the Wasserstein space $(\mathcal{P}_2(M), d_2^W)$ (see (4.2)). Moreover, each μ_t is absolutely continuous (Lemma 6.2). Recall (7.1) for the definition of the function $\beta_{K,N}^t$.

Definition 8.1 (N-Ricci curvature bounds, [LV1], [LV2]) For $K \in \mathbb{R}$ and $N \in (1,\infty]$, we say that (M, F, m) has N-*Ricci curvature bounded below by* K if, for any two absolutely continuous probability measures $\mu_0 = \rho_0 \mathsf{m}, \mu_1 = \rho_1 \mathsf{m} \in \mathcal{P}_c(M)$, $U \in \mathcal{DC}_N$ and for any $t \in (0,1)$, it holds that

$$U_{\mathsf{m}}(\mu_t) \leq (1-t) \int_{M\times M} \frac{\beta_{K,N}^{1-t}(d(x,y))}{\rho_0(x)} U\left(\frac{\rho_0(x)}{\beta_{K,N}^{1-t}(d(x,y))}\right) d\pi(x,y)$$
$$+ t \int_{M\times M} \frac{\beta_{K,N}^t(d(x,y))}{\rho_1(y)} U\left(\frac{\rho_1(y)}{\beta_{K,N}^t(d(x,y))}\right) d\pi(x,y), \qquad (8.1)$$

where $(\mu_t)_{t\in[0,1]}$ is the unique minimal geodesic from μ_0 to μ_1 and π is the unique optimal coupling of (μ_0, μ_1).

We usually require only the existence of a minimal geodesic $(\mu_t)_{t\in[0,1]}$ satisfying the above inequality in order to ensure the stability under the convergence of spaces. Thus we implicitly took advantage of the unique existence of minimal geodesics in the above definition. In addition, it is more consistent to include singular measures, whereas it is known by [LV1] that the above definition is sufficient to imply the corresponding generalized condition concerning possibly singular measures (see also [FV], [Vi2, Chapter 30]). Furthermore, as $\beta_{K,N}^t(r)$ is defined only for $r < \pi\sqrt{(N-1)/K}$ if $K > 0$, we apparently need to restrict ourselves to μ_0 and μ_1 such that $\mathrm{diam}(\mathrm{supp}\,\mu_0 \cup \mathrm{supp}\,\mu_1) \leq \pi\sqrt{(N-1)/K}$. Nonetheless, this local version is enough to imply a generalized Bonnet-Myers theorem $\mathrm{diam}\,M \leq \pi\sqrt{(N-1)/K}$, so that the restriction on $\mathrm{supp}\,\mu_0 \cup \mathrm{supp}\,\mu_1$ turns out unnecessary.

Sturm's *curvature-dimension condition* $\mathsf{CD}(K,N)$ uses the same inequality as Definition 8.1, but only for the Rényi and the relative entropies. These cases are indeed essential in the sense that an n-dimensional Riemannian manifold (M, g, vol_g) equipped with the Riemannian volume element satisfies $\mathsf{CD}(K,N)$ if and only if $\mathrm{Ric} \geq K$ and $n \leq N$. The K-convexity of the relative entropy considered in [vRS] corresponds to $\mathsf{CD}(K,\infty)$. Indeed, applying Definition 8.1 with $N = \infty$ to the relative entropy yields

$$\mathrm{Ent}_{\mathsf{m}}(\mu_t) \leq (1-t)\,\mathrm{Ent}_{\mathsf{m}}(\mu_0) + t\,\mathrm{Ent}_{\mathsf{m}}(\mu_1) - \frac{K}{2}(1-t)t d_2^W(\mu_0, \mu_1)^2. \qquad (8.2)$$

Now we give a proof of Theorem 1.2. This extends Sturm's and Lott and Villani's results concerning weighted Riemannian manifolds ([St1], [St2], [St3], [LV1], [LV2]).

8.2 Proof of Theorem 1.2

Take two absolutely continuous measures $\mu_0 = \rho_0 \mathsf{m}, \mu_1 = \rho_1 \mathsf{m} \in \mathcal{P}_c(M)$. Let ϕ be a c-concave function such that $\mathcal{F}_t(x) := \exp_x(\nabla(-t\phi)(x))$, $t \in [0,1]$, provides the unique

optimal transport map from μ_0 to μ_1 (Theorem 4.10). Then $\mu_t := (\mathcal{F}_t)_\#\mu_0$, $t \in [0,1]$, is the unique minimal geodesic from μ_0 to μ_1, and $\pi := (\mathrm{Id}_M \times \mathcal{F}_1)_\#\mu_0$ is the unique optimal coupling of (μ_0, μ_1). Fix $t \in (0,1)$. As μ_t is absolutely continuous, we can write $\mu_t = \rho_t\mathsf{m}$. We deduce from Theorem 5.2(iii) that

$$\rho_0(x) = \mathbf{J}_t(x)\rho_t\big(\mathcal{F}_t(x)\big) = \mathbf{J}_1(x)\rho_1\big(\mathcal{F}_1(x)\big) \tag{8.3}$$

holds for μ_0-a.e. x with $d\phi_x \neq 0$, where we put $\mathbf{J}_t(x) := \mathbf{D}[d(\mathcal{F}_t)_x]$ as in Proposition 5.3.

(i) Assume that (1.3) holds. We follow the proof of [St3, Theorem 1.7] using somewhat similar calculations to Lemma 7.2 and Theorem 7.3. Fix $x \in M$ at where $\rho_0(x) > 0$, $d\phi_x \neq 0$ and (8.3) holds. Put $v = \nabla(-\phi)(x)$ and let $\eta : [0,1] \longrightarrow M$ be the geodesic with $\dot\eta(0) = v$. Take an orthonormal basis $\{e_i\}_{i=1}^n$ of T_xM with respect to g_v such that $e_n = v/F(v)$. We extend them to $E_i(t) := d(\mathcal{F}_t)_x(e_i)$ for $t \in [0,1]$ and remark that every E_i is a Jacobi field, for $\mathcal{F}_t$ is a transport map along geodesics. We set $g_t = g_{\dot\eta(t)}$, $\mathcal{V}(t) = \mathcal{V}(\dot\eta(t)/F(v))$ and $E_i'(t) = D_{\dot\eta}^{\dot\eta}E_i(t)$ for brevity.

As in Lemma 7.2, we define $n \times n$ matrices $A(t)$ and $B(t)$ by $a_{ij}(t) = g_t(E_i, E_j)$ and $E_i'(t) = \sum_{j=1}^n b_{ij}(t)E_j(t)$. We remark that $\{E_i(t)\}_{i=1}^n$ is a basis of $T_{\eta(t)}M$ because $(\det A(t))^{1/2} = e^{\mathcal{V}(t)-\mathcal{V}(0)}\mathbf{J}_t(x) > 0$. We also define P_i as the parallel vector field along η (i.e., $D_{\dot\eta}^{\dot\eta}P_i \equiv 0$) with $P_i(0) = e_i$, and take the $n \times n$ matrix $C(t)$ with $P_i(t) = \sum_{j=1}^n c_{ij}(t)E_j(t)$. Note that $P_n(t) = \dot\eta(t)/F(v)$. Put $D = CBC^{-1}$ and recall that this is symmetric. Then $C' = -CB$ and calculations in Lemma 7.2 yield that

$$B' = -\operatorname{Ric}_{\dot\eta} A^{-1} - B^2, \qquad D' = CB'C^{-1} = -C\operatorname{Ric}_{\dot\eta} A^{-1}C^{-1} - D^2.$$

In particular, it follows from (2.3) that

$$d_{nn}' = -\sum_{i=1}^n d_{in}^2 \leq -d_{nn}^2. \tag{8.4}$$

Now we consider functions

$$h(t) := \big(e^{-\mathcal{V}(0)}\mathbf{J}_t(x)\big)^{1/N} = e^{-\mathcal{V}(t)/N}\big(\det A(t)\big)^{1/2N}, \qquad h_1(t) := e^{-\mathcal{V}(t)/(N-n)},$$

$$h_2(t) := \big(\det A(t)\big)^{1/2n}, \qquad f(t) := \log\big[\big(\det A(t)\big)^{1/2}\big] = n\log h_2(t).$$

Again we calculate as in Lemma 7.2 and observe $f' = \operatorname{trace}D$ and $f'' = -\operatorname{Ric}(\dot\eta) - \operatorname{trace}(D^2)$. We further define $\beta(t) := 1 + \int_0^t d_{nn}\,ds$ and $\alpha := f - \beta$. Then we see

$$\alpha'' = f'' - \beta'' = -\operatorname{Ric}(\dot\eta) - \operatorname{trace}(D^2) - d_{nn}' = -\operatorname{Ric}(\dot\eta) - \sum_{i,j=1}^n d_{ij}^2 + \sum_{i=1}^n d_{in}^2$$

$$\leq -\operatorname{Ric}(\dot\eta) - \sum_{i,j=1}^{n-1} d_{ij}^2 \leq -\operatorname{Ric}(\dot\eta) - \frac{1}{n-1}\bigg(\sum_{i=1}^{n-1} d_{ii}\bigg)^2 = -\operatorname{Ric}(\dot\eta) - \frac{(\alpha')^2}{n-1}.$$

Hence we have $[e^{\alpha/(n-1)}]''e^{-\alpha/(n-1)} \leq -\operatorname{Ric}(\dot\eta)/(n-1)$. We put $h_3 = h_1^{(N-n)/(N-1)}e^{\alpha/(N-1)}$ and obtain, as in Theorem 7.3,

$$h_3''h_3^{-1} \leq \frac{F(v)^2}{N-1}\bigg\{-\partial_{P_n}^2\mathcal{V} + \frac{(\partial_{P_n}\mathcal{V})^2}{N-n} - \operatorname{Ric}(P_n)\bigg\} \leq -\frac{F(v)^2K}{N-1}.$$

This implies that $\{h_3(t) - \mathbf{s}'_{K,N}(tr)h_3(0)\}/\mathbf{s}_{K,N}(tr)$ is nonincreasing in t, where we set $r = F(v) = d(x, \mathcal{F}_1(x))$. Therefore we find

$$h_3(t) \geq \frac{\mathbf{s}_{K,N}((1-t)r)}{\mathbf{s}_{K,N}(r)} h_3(0) + \frac{\mathbf{s}_{K,N}(tr)}{\mathbf{s}_{K,N}(r)} h_3(1). \tag{8.5}$$

On the other hand, (8.4) yields that $(e^\beta)'' \leq 0$, so that $e^{\beta(t)} \geq (1-t)e^{\beta(0)} + te^{\beta(1)}$. Combining this with (8.5) and the Hölder inequality shows that

$$h(t) = h_1(t)^{(N-n)/N} h_2(t)^{n/N} = h_1(t)^{(N-n)/N}(e^{f(t)})^{1/N} = h_3(t)^{(N-1)/N}(e^{\beta(t)})^{1/N}$$

$$\geq \{(1-t)e^{\beta(0)} + te^{\beta(1)}\}^{1/N} \left\{ \frac{\mathbf{s}_{K,N}((1-t)r)}{\mathbf{s}_{K,N}(r)} h_3(0) + \frac{\mathbf{s}_{K,N}(tr)}{\mathbf{s}_{K,N}(r)} h_3(1) \right\}^{(N-1)/N}$$

$$\geq \{(1-t)e^{\beta(0)}\}^{1/N} \left\{ \frac{\mathbf{s}_{K,N}((1-t)r)}{\mathbf{s}_{K,N}(r)} h_3(0) \right\}^{(N-1)/N} + \{te^{\beta(1)}\}^{1/N} \left\{ \frac{\mathbf{s}_{K,N}(tr)}{\mathbf{s}_{K,N}(r)} h_3(1) \right\}^{(N-1)/N}$$

$$= (1-t)\beta_{K,N}^{1-t}(r)^{1/N} h(0) + t\beta_{K,N}^t(r)^{1/N} h(1).$$

We consequently obtain

$$\mathbf{J}_t(x)^{1/N} \geq (1-t)\beta_{K,N}^{1-t}\big(d\big(x,\mathcal{F}_1(x)\big)\big)^{1/N} + t\beta_{K,N}^t\big(d\big(x,\mathcal{F}_1(x)\big)\big)^{1/N} \mathbf{J}_1(x)^{1/N}. \tag{8.6}$$

For any $U \in \mathcal{DC}_N$, as φ is nonincreasing and convex, we deduce from (8.3) and (8.6) that

$$U_{\mathsf{m}}(\mu_t) = \int_M U(\rho_t)\, d\mathsf{m} = \int_M \varphi(\rho_t^{-1/N})\rho_t\, d\mathsf{m}$$

$$= \int_M \varphi(\rho_t^{-1/N} \circ \mathcal{F}_t)\rho_t \circ \mathcal{F}_t \cdot \mathbf{J}_t\, d\mathsf{m} = \int_M \varphi\left(\frac{\mathbf{J}_t^{1/N}}{\rho_0^{1/N}} \right) d\mu_0$$

$$\leq \int_M \varphi\left((1-t)\frac{\beta_{K,N}^{1-t}(d(x,\mathcal{F}_1(x)))^{1/N}}{\rho_0(x)^{1/N}} + t\frac{\beta_{K,N}^t(d(x,\mathcal{F}_1(x)))^{1/N}}{\rho_1(\mathcal{F}_1(x))^{1/N}} \right) d\mu_0(x)$$

$$\leq (1-t)\int_M \varphi\left(\frac{\beta_{K,N}^{1-t}(d(x,\mathcal{F}_1(x)))^{1/N}}{\rho_0(x)^{1/N}} \right) d\mu_0(x) + t\int_M \varphi\left(\frac{\beta_{K,N}^t(d(x,\mathcal{F}_1(x)))^{1/N}}{\rho_1(\mathcal{F}_1(x))^{1/N}} \right) d\mu_0(x).$$

Plugging $\varphi(s) = s^N U(s^{-N})$ and $\pi = (\mathrm{Id}_M \times \mathcal{F}_1)_\# \mu_0$ into this shows the desired N-Ricci curvature bound. More precisely, we need to separately treat x with $d\phi_x = 0$, whereas then we find $\rho_0(x) = \rho_t(x) = \rho_1(x)$ by Lemma 6.1 and Theorem 5.2, and hence

$$\varphi\big(\rho_t(x)^{-1/N}\big) = \varphi\big((1-t)\rho_0(x)^{-1/N} + t\rho_1(x)^{-1/N}\big)$$
$$\leq (1-t)\varphi\big(\rho_0(x)^{-1/N}\big) + t\varphi\big(\rho_1(x)^{-1/N}\big).$$

To see the converse, assume that (M, F, m) has N-Ricci curvature bounded below by K. We shall modify a discussion in [St3, Theorem 1.7] taking [LV1] into accout. Fix a unit vector $v \in T_xM$ and put $a = (\partial_v \mathcal{V})/(N - n)$. Take the short geodesic $\eta : (-\delta, \delta) \longrightarrow M$ with $\dot{\eta}(0) = v$ and extend v to the C^∞-vector field $V := \nabla[d(\eta(-\delta), \cdot)]$ on $U := B^+(\eta(-\delta), 2\delta) \setminus \{\eta(-\delta)\}$. We introduce the Riemannian structure $g := g_V$ of U,

and remark that η is also a geodesic with respect to g ([Sh2, Lemma 6.2.1]). Moreover, $\mathrm{Ric}(v)$ with respect to F coincides with the Ricci curvature of v with respect to g ([Sh2, Proposition 6.2.2]).

Given $0 < \varepsilon \ll r \ll \delta$, we consider two open balls with respect to g,

$$A_+ := B^g\big(\eta(r), \varepsilon(1-ar)\big), \qquad A_- := B^g\big(\eta(-r), \varepsilon(1+ar)\big).$$

Note that

$$\mathsf{m}(A_+) = e^{-\mathcal{V}(\dot{\eta}(r))} c_n \varepsilon^n (1-ar)^n + O(\varepsilon^{n+1}), \quad \mathsf{m}(A_-) = e^{-\mathcal{V}(\dot{\eta}(-r))} c_n \varepsilon^n (1+ar)^n + O(\varepsilon^{n+1}),$$

where c_n denotes the volume of the standard unit ball in $\mathbb{R}^n$. Applying (8.1) with $U(r) = Nr(1-r^{-1/N})$ to $\mu_0 = \mathsf{m}(A_-)^{-1}\mathsf{m}|_{A_-}$, $\mu_1 = \mathsf{m}(A_+)^{-1}\mathsf{m}|_{A_+}$ and $t = 1/2$ shows the following Brunn-Minkowski inequality ([St3, Proposition 2.1], Corollary 9.4(ii))

$$\mathsf{m}\big(Z_{1/2}(A_-, A_+)\big)^{1/N} \geq \frac{1}{2}\beta_{K,N}^{1/2}\big(2r + O(\varepsilon)\big)^{1/N}\{\mathsf{m}(A_-)^{1/N} + \mathsf{m}(A_+)^{1/N}\}.$$

We also observe

$$\beta_{K,N}^{1/2}(2r) = \left(\frac{2\mathsf{s}_{K,N}(r)}{\mathsf{s}_{K,N}(2r)}\right)^{N-1} = 1 + \frac{K}{2}r^2 + O(r^4).$$

In addition, we deduce from $a = (\partial_v \mathcal{V})/(N-n)$ that

$$\frac{\partial^2}{\partial r^2}\big[e^{-\mathcal{V}(\dot{\eta}(r))/N}(1-ar)^{n/N}\big]\Big|_{r=0}$$
$$= \left\{\frac{n}{N}\left(\frac{n}{N}-1\right)a^2 + \frac{2n}{N^2}a\partial_v\mathcal{V} + \frac{1}{N^2}(\partial_v\mathcal{V})^2 - \frac{1}{N}\partial_v^2\mathcal{V}\right\}e^{-\mathcal{V}(v)/N}$$
$$= \left\{\frac{(\partial_v\mathcal{V})^2}{N(N-n)} - \frac{1}{N}\partial_v^2\mathcal{V}\right\}e^{-\mathcal{V}(v)/N}.$$

Thus it follows that

$$\frac{\mathsf{m}(Z_{1/2}(A_-, A_+))}{c_n\varepsilon^n}$$
$$\geq \frac{1}{2^N}\left(1 + \frac{K}{2}r^2 + O(r^4)\right)\left\{2e^{-\mathcal{V}(v)/N} + \left(\frac{(\partial_v\mathcal{V})^2}{N(N-n)} - \frac{\partial_v^2\mathcal{V}}{N}\right)e^{-\mathcal{V}(v)/N}r^2 + O(r^4)\right\}^N$$
$$= \left(1 + \frac{K}{2}r^2\right)e^{-\mathcal{V}(v)}\left\{1 + \frac{1}{2}\left(\frac{(\partial_v\mathcal{V})^2}{N-n} - \partial_v^2\mathcal{V}\right)r^2\right\} + O(r^4)$$
$$= e^{-\mathcal{V}(v)}\left\{1 + \frac{1}{2}\left(K - \partial_v^2\mathcal{V} + \frac{(\partial_v\mathcal{V})^2}{N-n}\right)r^2\right\} + O(r^4).$$

On the other hand, if we choose an orthonormal basis $\{e_i\}_{i=1}^n$ of $T_x M$ with respect to g with $e_n = v$, and denote by k_i the sectional curvature (with respect to g) of the plane spanned by e_i and e_n, then

$$\frac{\mathsf{m}(Z_{1/2}(A_-, A_+))}{c_n\varepsilon^n} = e^{-\mathcal{V}(v)}\prod_{i=1}^{n-1}\left(1 + \frac{k_i}{2}r^2 + O(r^3)\right) = e^{-\mathcal{V}(v)}\left(1 + \frac{\mathrm{Ric}(v)}{2}r^2\right) + O(r^3).$$

Therefore we conclude that $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} - (\partial_v \mathcal{V})^2/(N-n) \geq K$.

(ii) Assume $\mathrm{Ric}(v) \geq K$ and $\partial_v \mathcal{V} = 0$. In this case, (8.6) with $N = n$ directly follows from Proposition 5.3 and Theorem 7.1 with $H = 0$, and then we discuss as in (i).

If (M, F, m) has n-Ricci curvature bounded below by K, then it also has N-Ricci curvature bounded below by K for any $N \in (n, \infty)$. Hence (i) yields $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} - (\partial_v \mathcal{V})^2/(N-n) \geq K$. Letting N tend to n shows that $\partial_v \mathcal{V} \equiv 0$ and then $\mathrm{Ric}(v) \geq K$.

(iii) It costs no generality to assume that (M, F) is compact (because we deal with only compactly supported measures in Definition 8.1). If $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} \geq K$, then we have

$$\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} - \frac{(\partial_v \mathcal{V})^2}{N - n} \geq K - \frac{1}{N - n} \sup_{v \in TM, F(v)=1} (\partial_v \mathcal{V})^2$$

for any $N \in (n, \infty)$. Thus (i) shows that (M, F, m) has ∞-Ricci curvature bounded below by $K - \sup_{v \in TM, F(v)=1} (\partial_v \mathcal{V})^2/(N - n)$, and letting N diverge to infinity yields the desired curvature bound K.

The direction from ∞-Ricci curvature bounded below by K to $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} \geq K$ is the special case ($a = 0$) of the latter half of the proof of (i) (see [vR, Theorem 1.1]). Indeed, (8.2) applied to $\mu_0 = \mathsf{m}(A_-)^{-1}\mathsf{m}|_{A_-}$, $\mu_1 = \mathsf{m}(A_+)^{-1}\mathsf{m}|_{A_+}$ and $t = 1/2$ yields (see Corollary 9.5(ii))

$$\log \mathsf{m}\big(Z_{1/2}(A_-, A_+)\big) \geq \frac{1}{2} \log \mathsf{m}(A_-) + \frac{1}{2} \log \mathsf{m}(A_+) + \frac{K}{8} d_2^W(\mu_0, \mu_1)^2.$$

Hence we have

$$\frac{\mathsf{m}(Z_{1/2}(A_-, A_+))}{c_n \varepsilon^n} \geq e^{-\{\mathcal{V}(\dot{\eta}(-r)) + \mathcal{V}(\dot{\eta}(r))\}/2} e^{Kr^2/2} + O(\varepsilon)$$

$$= e^{-\mathcal{V}(v)}\left(1 - \frac{\partial_v^2 \mathcal{V}}{2} r^2\right)\left(1 + \frac{K}{2} r^2\right) + O(r^4) = e^{-\mathcal{V}(v)}\left\{1 + \frac{K - \partial_v^2 \mathcal{V}}{2} r^2\right\} + O(r^4).$$

Therefore we obtain $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} \geq K$. □

8.3 Remarks and corollaries

We close this section with selected remarks and corollaries.

Remark 8.2 (a) Let us briefly comment on the reason why the same characterization as Theorem 1.2(i) does not work for the measure contraction property (i.e., the converse of Theorem 7.3 fails, compare with [Oh1], [St3]). In the proof of Theorem 1.2(i), we considered a part of a cone bounded by A_+ and A_-. If one could find the origin of this cone (in $B^+(x, \pi\sqrt{(N-1)/K})$ or $B^-(x, \pi\sqrt{(N-1)/K})$ if $K > 0$), then the measure contraction property is applied. The origin is, however, $\eta(a^{-1})$ by construction, so that we need the infinite extendability of minimal geodesics even if $K \leq 0$, and $\eta(a^{-1})$ is nonsense typically when $K > 0$ and $\partial_v \mathcal{V} = 0$. We actually know that a sufficiently small ball in $\mathbb{R}^n$ has positive curvature in terms of the measure contraction property with $N > n$ ([St3, Remark 5.6]).

(b) The most restricted situation (ii) in Theorem 1.2 still admits a number of non-Riemannian spaces. As we mentioned at the end of Section 7, if (M, F) is Berwald

type, then the Busemann-Hausdorff measure satisfies $\partial_v \mathcal{V} \equiv 0$. In particular, Minkowski spaces have nonnegative n-Ricci curvature (see [Vi2, Theorem in page 908] for Cordero-Erausquin's preceding work on Banach spaces).

It is possible to use (8.6) instead of Proposition 5.3 in the proof of Theorem 1.1, and then we obtain the following variant. We remark that this corollary does not follow from the combination of Theorems 1.1 and 7.3 (see also Remark 8.2(a) above).

Corollary 8.3 *Assume that $n \geq 2$ and there are constants $K \in \mathbb{R}$ and $N \in (n, \infty)$ such that $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} - (\partial_v \mathcal{V})^2 / (N - n) \geq K$ holds for all unit vectors $v \in TM$. Take three nonnegative measurable functions $f, g, h : M \longrightarrow [0, \infty)$ and measurable sets $A, B \subset M$ with $\int_A f \, d\mathsf{m} = \int_B g \, d\mathsf{m} = 1$. If there is $t \in (0, 1)$ such that*

$$\frac{1}{h(z)^{1/N}} \leq (1 - t) \left(\frac{\beta_{K,N}^{1-t}(d(x, y))}{f(x)} \right)^{1/N} + t \left(\frac{\beta_{K,N}^{t}(d(x, y))}{g(y)} \right)^{1/N}$$

holds for all $x \in A$, $y \in B$ and $z \in Z_t(x, y)$, then we have $\int_M h \, d\mathsf{m} \geq 1$.

As we have already mentioned, applying Theorem 1.2(iii) to the relative entropy yields the K-convexity (8.2). Moreover, Theorem 1.2 contains the following weighted version. Take a C^∞-function $V : M \longrightarrow \mathbb{R}$ and suppose that $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} + \partial_v^2 V \geq K$ holds for every unit vector $v \in TM$, where we put $\partial_v^2 V = (V \circ \eta)''(0)$ using the geodesic $\eta : (-\varepsilon, \varepsilon) \longrightarrow M$ with $\dot{\eta}(0) = v$. Then $(M, F, e^{-V}\mathsf{m})$ has ∞-Ricci curvature bounded below by K. In particular, the corresponding *free energy* $\mathrm{Ent}_{e^{-V}\mathsf{m}}(\rho\mathsf{m}) = \mathrm{Ent}_{\mathsf{m}}(\rho\mathsf{m}) + \int_M V \rho \, d\mathsf{m}$ is K-convex. The K-convexity of the relative entropy as well as the free energy plays an essential role in the investigation of gradient flows on Wasserstein spaces. It has been developed in [JKO], [AGS], [Oh3] and [Sa] in the cases where underlying spaces are Euclidean spaces, Hilbert spaces, Riemannian manifolds and Alexandrov spaces. Meanwhile, in the Finsler setting, even gradient flows on Finsler manifolds are not yet studied well.

We finally recall functional inequalities following from Theorem 1.2, they are related to concentration of measures studied in the next section.

Corollary 8.4 ([LV1], see also [CMS2]) *Let (M, F, m) be a compact Finsler manifold satisfying $n \geq 2$, $\mathsf{m}(M) = 1$ and $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} \geq K$ for some $K \in \mathbb{R}$ and all unit vectors $v \in TM$. Then the following hold.*

(i) *(Talagrand inequality/Transport cost inequality) For any absolutely continuous measure $\mu \in \mathcal{P}(M)$, we have*

$$\mathrm{Ent}_{\mathsf{m}}(\mu) \geq \frac{K}{2} \max \left\{ d_2^W(\mathsf{m}, \mu)^2, d_2^W(\mu, \mathsf{m})^2 \right\}.$$

(ii) *(Logarithmic Sobolev inequality) For any Lipschitz continuous function $f : M \longrightarrow \mathbb{R}$ with $\int_M f^2 \, d\mathsf{m} = 1$, we have*

$$\mathrm{Ent}_{\mathsf{m}}(\mu) \leq 2 \left(\int_M F(\nabla|f|)^2 \, d\mathsf{m} \right)^{1/2} d_2^W(\mathsf{m}, \mu) - \frac{K}{2} d_2^W(\mathsf{m}, \mu)^2,$$

34

where we set $\mu = f^2 \mathsf{m}$. In particular, if $K > 0$, then it holds that

$$\mathrm{Ent}_{\mathsf{m}}(\mu) \leq \frac{2}{K} \int_M F(\nabla f)^2 \, d\mathsf{m}.$$

(iii) (Global Poincaré inequality) *If $K > 0$, then we have, for any Lipschitz continuous function $f : M \longrightarrow \mathbb{R}$ with $\int_M f \, d\mathsf{m} = 0$,*

$$\int_M f^2 \, d\mathsf{m} \leq \frac{1}{K} \int_M F(\nabla f)^2 \, d\mathsf{m}.$$

Proof. The proof is similar to [LV1, §6] by taking $\nabla(f^2) = 2|f|\nabla|f|$ into account. Here we mention only the reason why $F(\nabla|f|)$ can be replaced with $F(\nabla f)$ in the latter half of (ii). We first remark that the curvature bound $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} \geq K$ is common between F and its reverse $\bar{F}$, and hence we have

$$\mathrm{Ent}_{\mathsf{m}}(\mu) \leq \frac{2}{K} \int_M F(\nabla|f|)^2 \, d\mathsf{m}, \quad \mathrm{Ent}_{\mathsf{m}}(\mu) \leq \frac{2}{K} \int_M F\big(\nabla(-|f|)\big)^2 \, d\mathsf{m}.$$

Note that, if $\bar{\nabla}|f|$ stands for the gradient vector with respect to $\bar{F}$, then $\bar{F}(\bar{\nabla}|f|) = \bar{F}(-\nabla(-|f|)) = F(\nabla(-|f|))$. Therefore we are done if $f \geq 0$ or $f \leq 0$. Otherwise we divide f into $f_+ := \max\{f, 0\}$ and $f_- := \max\{-f, 0\}$, and apply the above inequalities to $f_+/(\int_M f_+^2 \, d\mathsf{m})^{1/2}$ and $f_-/(\int_M f_-^2 \, d\mathsf{m})^{1/2}$, respectively. $\qquad\square$

Corollary 8.5 (Lichnerowicz inequality, [LV2, Theorem 5.34]) *Suppose that $n \geq 2$ and there are constants $K > 0$ and $N \in (n, \infty)$ such that $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} - (\partial_v \mathcal{V})^2/(N-n) \geq K$ holds for all unit vectors $v \in TM$. Then we have, for any Lipschitz continuous function $f : M \longrightarrow \mathbb{R}$ with $\int_M f \, d\mathsf{m} = 0$,*

$$\int_M f^2 \, d\mathsf{m} \leq \frac{N-1}{KN} \int_M F(\nabla f)^2 \, d\mathsf{m}.$$

9 Applications

This final section is devoted to applications of the Finsler Borell-Brascamp-Lieb inequality (Theorem 1.1, Corollary 8.3) and the volume comparison theorems (Theorems 7.1, 7.3).

9.1 Further interpolation inequalities

We first treat interpolation inequalities closely related to the Borell-Brascamp-Lieb inequality. See [Ga] and [CMS1] for their correlation and historical background.

For $p \in \mathbb{R} \setminus 0$, $t \in (0, 1)$ and $a, b \in [0, \infty)$, we define

$$\mathcal{M}_t^p(a, b) := \{(1-t)a^p + tb^p\}^{1/p}.$$

We set $\mathcal{M}_t^p(a, b) := 0$ if $p < 0$ and $ab = 0$. We also define, as the limits,

$$\mathcal{M}_t^0(a, b) := a^{1-t}b^t, \quad \mathcal{M}_t^\infty(a, b) := \max\{a, b\}, \quad \mathcal{M}_t^{-\infty}(a, b) := \min\{a, b\}.$$

By convention, we set $\mathcal{M}_t^p(0,\infty) := 0$ if $p \in [-\infty, 0]$, $\mathcal{M}_t^p(0,\infty) := \infty$ if $p \in (0,\infty]$, and $\mathcal{M}_t^p(\infty,\infty) := \infty$ for $p \in [-\infty,\infty]$.

We will use the following type of Hölder inequality. For $a,b,c,d \in [0,\infty)$, $p,q \in (-\infty,\infty]$ and $r \in [-\infty,\infty]$ with $p + q \geq 0$ as well as $p^{-1} + q^{-1} = r^{-1}$, it holds that

$$\mathcal{M}_t^r(ac, bd) \leq \mathcal{M}_t^p(a,b)\mathcal{M}_t^q(c,d). \tag{9.1}$$

Here $r = -\infty$ if $p = -q \neq 0$, $r = 0$ if $p = 0$ or $q = 0$, and $r = \infty$ if $p = q = \infty$. If $p,q \in (0,\infty)$, then this is the usual Hölder inequality. By taking the limits, we obtain (9.1) for $p,q,r \in [0,\infty]$. If $0 < -q < p < \infty$ and $a,b,c,d \in (0,\infty)$, then $r \in (-\infty, q)$ and we see that $\mathcal{M}_t^{-q}(c^{-1}, d^{-1}) \leq \mathcal{M}_t^p(a,b)\mathcal{M}_t^{-r}((ac)^{-1}, (bd)^{-1})$. This is indeed (9.1) by taking $\mathcal{M}_t^q(c,d) = \mathcal{M}_t^{-q}(c^{-1}, d^{-1})^{-1}$ into account. Taking the limits shows (9.1) for $0 < -q < p < \infty$ and $a,b,c,d \in [0,\infty)$. Finally, taking the limits again yields (9.1) for $0 < -q \leq p \leq \infty$ and $r \in [-\infty, q]$.

Corollary 9.1 (*p-mean inequality*) *Take* $t \in (0,1)$, $p \in [-1/n, \infty]$, *three nonnegative measurable functions* $f, g, h : M \longrightarrow [0,\infty)$ *and nonempty measurable sets* $A, B \subset M$ *with* $\int_A f\, d\mathsf{m} = \int_M f\, d\mathsf{m}$ *and* $\int_B g\, d\mathsf{m} = \int_M g\, d\mathsf{m}$. *If we have*

$$h(z) \geq \mathcal{M}_t^p\left(\frac{f(x)}{\mathfrak{v}_t^{>}(x,y)}, \frac{g(y)}{\mathfrak{v}_t^{<}(x,y)}\right)$$

for all $x \in A$, $y \in B$ *and* $z \in Z_t(x,y)$, *then* $\int_M h\, d\mathsf{m} \geq \mathcal{M}_t^{p/(1+np)}(\int_M f\, d\mathsf{m}, \int_M g\, d\mathsf{m})$ *holds. Here* $p/(1+np) = -\infty$ *if* $p = -1/n$ *and* $p/(1+np) = 1/n$ *if* $p = \infty$.

Proof. If $\|f\|_1, \|g\|_1 \in (0,\infty)$, then we put $\hat{h} = \mathcal{M}_t^{p/(1+np)}(\|f\|_1, \|g\|_1)^{-1} h$ and shall see that $(f/\|f\|_1, g/\|g\|_1, \hat{h})$ satisfies the assumption (1.1) of Theorem 1.1. Fix $x \in A$ and $y \in B$. If $f(x)g(y) = 0$, then (1.1) is clearly satisfied. In case of $f(x)g(y) > 0$, we deduce from the hypothesis on h and (9.1) that, for any $z \in Z_t(x,y)$,

$$\hat{h}(z) \geq \mathcal{M}_t^{p/(1+np)}(\|f\|_1, \|g\|_1)^{-1}\mathcal{M}_t^p\left(\frac{f(x)}{\mathfrak{v}_t^{>}(x,y)}, \frac{g(y)}{\mathfrak{v}_t^{<}(x,y)}\right)$$

$$\geq \mathcal{M}_t^{1/n}\left(\frac{\mathfrak{v}_t^{>}(x,y)\|f\|_1}{f(x)}, \frac{\mathfrak{v}_t^{<}(x,y)\|g\|_1}{g(y)}\right)^{-1}.$$

Thus Theorem 1.1 yields $\|\hat{h}\|_1 \geq 1$ as required.

If $\|f\|_1 = 0$, then the conclusion is obvious ($\|h\|_1 \geq 0$) unless $\|g\|_1 > 0$ and $p > 0$. For $p \in (0,\infty]$, the assumption on h means that $h(z) \geq t^{1/p}g(y)/\mathfrak{v}_t^{<}(x,y)$. Fix $x \in A$ and define the map $\tau_t^{<} : B \setminus \text{Cut}(x) \longrightarrow M$ by $\tau_t^{<}(y) := \eta(t)$, where $\eta : [0,1] \longrightarrow M$ is the unique minimal geodesic from x to y. Then $\tau_t^{<}(y) \in Z_t(x,y)$ and we have

$$\|h\|_1 \geq \int_{\tau_t^{<}(B \setminus \text{Cut}(x))} h\, d\mathsf{m} = \int_B (h \circ \tau_t^{<})\mathbf{D}[d\tau_t^{<}]\, d\mathsf{m} \geq \int_B \frac{t^{1/p}g(y)}{\mathfrak{v}_t^{<}(x,y)}t^n\mathfrak{v}_t^{<}(x,y)\, d\mathsf{m}(y)$$

$$= t^{(1+np)/p}\int_B g\, d\mathsf{m} = \mathcal{M}_t^{p/(1+np)}(0, \|g\|_1).$$

The case $\|f\|_1 + \|g\|_1 = \infty$ is deduced by approximating f and g from below. $\qquad\square$

The least case $p = -1/n$ corresponds to Theorem 1.1 and we have seen that it is the strongest. As the particular case of $p = 0$, we obtain the following.

Corollary 9.2 (Prékopa-Leindler inequality) *Take $t \in (0,1)$, three nonnegative measurable functions $f, g, h : M \longrightarrow [0, \infty)$ and nonempty measurable sets $A, B \subset M$ with $\int_A f \, d\mathsf{m} = \int_M f \, d\mathsf{m}$ and $\int_B g \, d\mathsf{m} = \int_M g \, d\mathsf{m}$. If we have*

$$h(z) \geq \left(\frac{f(x)}{\mathfrak{v}_t^{>}(x,y)} \right)^{1-t} \left(\frac{g(y)}{\mathfrak{v}_t^{<}(x,y)} \right)^{t}$$

for all $x \in A$, $y \in B$ and $z \in Z_t(x,y)$, then $\int_M h \, d\mathsf{m} \geq (\int_M f \, d\mathsf{m})^{1-t}(\int_M g \, d\mathsf{m})^{t}$ holds.

Corollary 9.3 (Brunn-Minkowski inequality) *For any nonempty measurable sets $A, B \subset M$ and $t \in (0,1)$, we have*

$$\mathsf{m}\big(Z_t(A,B)\big)^{1/n} \geq (1-t) \inf_{x \in A, y \in B} \mathfrak{v}_t^{>}(x,y)^{1/n} \cdot \mathsf{m}(A)^{1/n} + t \inf_{x \in A, y \in B} \mathfrak{v}_t^{<}(x,y)^{1/n} \cdot \mathsf{m}(B)^{1/n}.$$

Proof. If $\mathsf{m}(A), \mathsf{m}(B) \in (0, \infty)$, then we get this by applying Theorem 1.1 to normalized characteristic functions $f = \mathsf{m}(A)^{-1}\mathbf{1}_A$, $g = \mathsf{m}(B)^{-1}\mathbf{1}_B$ and

$$h = \left[(1-t)\{ \inf_{x \in A, y \in B} \mathfrak{v}_t^{>}(x,y) \cdot \mathsf{m}(A) \}^{1/n} + t\{ \inf_{x \in A, y \in B} \mathfrak{v}_t^{<}(x,y) \cdot \mathsf{m}(B) \}^{1/n} \right]^{-n} \mathbf{1}_{Z_t(A,B)}.$$

If $\mathsf{m}(A) = 0$ and $\mathsf{m}(B) \in (0, \infty)$, then we argue as in the proof of Corollary 9.1. The case $\mathsf{m}(A) = \mathsf{m}(B) = 0$ is trivial and the case $\mathsf{m}(A) + \mathsf{m}(B) = \infty$ follows by approximating A and B by increasing subsets. $\square$

We can derive variants of the above inequalities using Corollary 8.3 (and Theorem 7.3) instead of Theorem 1.1.

Corollary 9.4 *Suppose that $n \geq 2$ and there are constants $K \in \mathbb{R}$ and $N \in (n, \infty)$ such that $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} - (\partial_v \mathcal{V})^2/(N-n) \geq K$ holds for all unit vectors $v \in TM$. Then the following hold.*

(i) *Take $t \in (0,1)$, $p \in [-1/N, \infty]$, three nonnegative measurable functions $f, g, h : M \longrightarrow [0, \infty)$ and nonempty measurable sets $A, B \subset M$ with $\int_A f \, d\mathsf{m} = \int_M f \, d\mathsf{m}$ and $\int_B g \, d\mathsf{m} = \int_M g \, d\mathsf{m}$. If we have*

$$h(z) \geq \mathcal{M}_t^p \left(\frac{f(x)}{\beta_{K,N}^{1-t}(d(x,y))}, \frac{g(y)}{\beta_{K,N}^{t}(d(x,y))} \right)$$

for all $x \in A$, $y \in B$ and $z \in Z_t(x,y)$, then $\int_M h \, d\mathsf{m} \geq \mathcal{M}_t^{p/(1+Np)}(\int_M f \, d\mathsf{m}, \int_M g \, d\mathsf{m})$ holds. Here $p/(1+Np) = -\infty$ if $p = -1/N$ and $p/(1+Np) = 1/N$ if $p = \infty$.

(ii) *For any nonempty measurable sets $A, B \subset M$ and $t \in (0,1)$, we have*

$$\mathsf{m}\big(Z_t(A,B)\big)^{1/N} \geq (1-t) \inf_{x \in A, y \in B} \beta_{K,N}^{1-t}\big(d(x,y)\big)^{1/N} \cdot \mathsf{m}(A)^{1/N}$$

$$+ t \inf_{x \in A, y \in B} \beta_{K,N}^{t}\big(d(x,y)\big)^{1/N} \cdot \mathsf{m}(B)^{1/N}.$$

We further obtain the following infinite dimensional version by letting N diverge to infinity in (i) above, and by (8.2) along with Jensen's inequality.

Corollary 9.5 *Suppose that $n \geq 2$ and there is a constant $K \in \mathbb{R}$ such that $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} \geq K$ holds for all unit vectors $v \in TM$. Then the following hold.*

(i) *Take $t \in (0,1)$, $p \in [0, \infty]$, three nonnegative measurable functions $f, g, h : M \longrightarrow [0, \infty)$ and nonempty measurable sets $A, B \subset M$ with $\int_A f \, d\mathsf{m} = \int_M f \, d\mathsf{m}$ and $\int_B g \, d\mathsf{m} = \int_M g \, d\mathsf{m}$. If we have*

$$h(z) \geq \mathcal{M}_t^p \left(\frac{f(x)}{\beta_{K,\infty}^{1-t}(d(x,y))}, \frac{g(y)}{\beta_{K,\infty}^t(d(x,y))} \right)$$

for all $x \in A$, $y \in B$ and $z \in Z_t(x,y)$, then $\int_M h \, d\mathsf{m} \geq (\int_M f \, d\mathsf{m})^{1-t}(\int_M g \, d\mathsf{m})^t$ holds.

(ii) *For any bounded measurable sets $A, B \subset M$ with $\mathsf{m}(A), \mathsf{m}(B) > 0$ and $t \in (0,1)$, we have*

$$\log \mathsf{m}\big(Z_t(A,B)\big) \geq (1-t) \log \mathsf{m}(A) + t \log \mathsf{m}(B) + \frac{K}{2}(1-t)t d_2^W(\mu, \nu)^2,$$

where $\mu = \mathsf{m}(A)^{-1} \mathsf{m}|_A$ and $\nu = \mathsf{m}(B)^{-1} \mathsf{m}|_B$.

9.2 Concentration of measures

We finally discuss the concentration of measure phenomenon along the lines of [Le, §1.6] (as well as [CMS1], [CMS2]). See [Sh2, §4] for related work. In this topic, it turns out that both $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} \geq K$ and $\mathrm{Ric}(v) \geq K$ with $\partial_v \mathcal{V} \geq -H$ give the same rate of concentration.

We suppose that $\mathrm{Ric}(v) \geq K > 0$ and $\partial_v \mathcal{V} \geq -H$ hold for every unit vector $v \in TM$. Given a function $f : M \longrightarrow (-\infty, \infty]$ which is not identically ∞, define

$$Qf(y) := \inf_{x \in M} \left\{ f(x) + \frac{K}{4} d(x,y)^2 - H d(x,y) \right\}.$$

Then it follows from (7.2) and Theorem 7.1 with $t = 1/2$ that, for any $x, y \in M$ with $f(x) < \infty$,

$$e^{-f(x)} e^{Qf(y)} \leq e^{Kd(x,y)^2/4} e^{-Hd(x,y)} \leq e^{-Hd(x,y)} \beta_{K,n}^{1/2}\big(d(x,y)\big)^2 \leq \mathfrak{v}_{1/2}^>(x,y) \mathfrak{v}_{1/2}^<(x,y).$$

Hence we can apply Corollary 9.2 to $(f, g, h) = (e^{-f}, e^{Qf}, 1)$ and $t = 1/2$ to obtain the *infimum-convolution inequality*

$$\int_M e^{-f} \, d\mathsf{m} \int_M e^{Qf} \, d\mathsf{m} \leq \mathsf{m}(M)^2. \tag{9.2}$$

Proposition 9.6 *If $n \geq 2$ and there are constants $K > 0$ and $H \geq 0$ such that $\mathrm{Ric}(v) \geq K$ and $\partial_v \mathcal{V} \geq -H$ hold for all unit vectors $v \in TM$, then we have the following:*

(i) *For any measurable set $A \subset M$ with $\mathsf{m}(A) > 0$ and $r > 0$, it holds that*

$$\frac{1}{\mathsf{m}(M)}\mathsf{m}\big(\{y \in M \mid d(A, y) \geq r\}\big) \leq \frac{\mathsf{m}(M)}{\mathsf{m}(A)}e^{2H^2/K}e^{-Kr^2/8}.$$

(ii) *For any 1-Lipschitz function $f : M \longrightarrow \mathbb{R}$ and $r > 0$, it holds that*

$$\frac{1}{\mathsf{m}(M)}\mathsf{m}\left(\left\{x \in M \,\middle|\, f(x) \geq \mathsf{m}(M)^{-1}\int_M f \, d\mathsf{m} + r\right\}\right) \leq e^{2H^2/K}e^{-Kr^2/8}.$$

(iii) *If we assume $\mathrm{Ric}(v) + \partial_v^2 \mathcal{V} \geq K > 0$ instead of $\mathrm{Ric}(v) \geq K$ and $\partial_v \mathcal{V} \geq -H$, then we can replace $e^{2H^2/K}e^{-Kr^2/8}$ with $e^{-Kr^2/4}$ in both (i) and (ii).*

Proof. (i) Define a function f on M by 0 on A and ∞ on $M \setminus A$. Then we find, if $d(A, y) \geq r$,

$$Qf(y) = \inf_{x \in A}\left\{\frac{K}{4}d(x, y)^2 - Hd(x, y)\right\} \geq \frac{K}{8}\inf_{x \in A}d(x, y)^2 + \frac{K}{8}\inf_{x \in A}\left\{d(x, y)^2 - \frac{8H}{K}d(x, y)\right\}$$

$$\geq \frac{K}{8}r^2 - \frac{K}{8}\left(\frac{4H}{K}\right)^2 = \frac{K}{8}r^2 - \frac{2H^2}{K}.$$

Thus applying (9.2) to f after Chebyshev's inequality yields

$$\mathsf{m}\big(\{y \in M \mid d(A, y) \geq r\}\big) \leq e^{-Kr^2/8 + 2H^2/K}\int_M e^{Qf} \, d\mathsf{m} \leq \frac{\mathsf{m}(M)^2}{\mathsf{m}(A)}e^{2H^2/K}e^{-Kr^2/8}.$$

(ii) We first remark that the assertion is obvious if $r \leq 4H/K$, so that we assume $r > 4H/K$. For simplicity, we put $a := \mathsf{m}(M)$ in the following. Jensen's inequality combined with (9.2) shows

$$\int_M e^{Qf} \, d\mathsf{m} \leq a^2 \left(\int_M e^{-f} \, d\mathsf{m}\right)^{-1} \leq ae^{a^{-1}\int_M f \, d\mathsf{m}}.$$

If f is λ-Lipschitz, then we have, for any $y \in M$,

$$Qf(y) \geq f(y) + \inf_{x \in M}\left\{-\lambda d(x, y) + \frac{K}{4}d(x, y)^2 - Hd(x, y)\right\}$$

$$= f(y) + \frac{K}{4}\inf_{x \in M}\left\{d(x, y)^2 - \frac{4(\lambda + H)}{K}d(x, y)\right\}$$

$$\geq f(y) - \frac{K}{4}\frac{4(\lambda + H)^2}{K^2} = f(y) - \frac{(\lambda + H)^2}{K}.$$

Hence we have

$$\int_M e^f \, d\mathsf{m} \leq e^{(\lambda + H)^2/K}\int_M e^{Qf} \, d\mathsf{m} \leq ae^{(\lambda + H)^2/K}e^{a^{-1}\int_M f \, d\mathsf{m}}.$$

This implies, for a 1-Lipschitz function f and any $\lambda > 0$,

$$\int_M e^{\lambda f - a^{-1} \int_M \lambda f \, d\mathsf{m}} \, d\mathsf{m} \leq a e^{(\lambda + H)^2 / K}.$$

Therefore we obtain

$$\mathsf{m}\left(\left\{x \in M \,\Big|\, f(x) \geq a^{-1} \int_M f \, d\mathsf{m} + r\right\}\right)$$

$$\leq \inf_{\lambda > 0} e^{-\lambda r} \int_M e^{\lambda f - a^{-1} \int_M \lambda f \, d\mathsf{m}} \, d\mathsf{m} \leq a \inf_{\lambda > 0} e^{-\lambda r + (\lambda + H)^2 / K}.$$

Choosing $\lambda = (2 + \sqrt{2})(Kr - 2\sqrt{2}H)/4 > 0$ completes the proof.

(iii) We know that $\operatorname{Ric}(v) + \partial_v^2 \mathcal{V} \geq K > 0$ ensures $\mathsf{m}(M) < \infty$ by [St2, Theorem 4.26]. Then a similar discussion to (i) and (ii) using Corollary 9.5(i) and

$$Qf(y) := \inf_{x \in M}\left\{f(x) + \frac{K}{4}d(x,y)^2\right\}$$

enables us to replace $e^{2H^2/K}e^{-Kr^2/8}$ with $e^{-Kr^2/4}$ in both (i) and (ii). □

Note that (iii) above recovers Riemannian normal concentration in [GM] derived from the Lévy-Gromov isoperimetric inequality (see [Le] for more information).

As an immediate corollary to Proposition 9.6, we find a new example of Lévy family. For a (symmetric or nonsymmetric) metric space (X, d) equipped with a Borel probability measure μ on X, we define the *concentration function* by

$$\alpha_{(X,d,\mu)}(r) := \sup\left\{1 - \mu\big(B^+(A, r)\big) \,\big|\, A \subset X, \ \mu(A) \geq 1/2\right\}$$

for $r > 0$, where $B^+(A, r) := \{y \in X \mid d(A, y) < r\}$. Then a family of metric measure spaces $\{(X_k, d_k, \mu_k)\}_{k \in \mathbb{N}}$ is called a *Lévy family* if we have

$$\lim_{k \to \infty} \alpha_{(X_k, d_k, \mu_k)}(\operatorname{diam} X_k \cdot r) = 0$$

for all $r > 0$. This concept was introduced in [GM] and has rich geometric and topological applications (see [Le] and the references therein).

Corollary 9.7 *Let $\{(M_k, F_k, \mu_k)\}_{k \in \mathbb{N}}$ be a sequence of Finsler manifolds of finite total volume equipped with normalized measures $\mu_k = \mathsf{m}_k(M_k)^{-1}\mathsf{m}_k$. Assume that $\operatorname{diam} M_k \geq 1$ and that either of the following holds.*

(a) *There are constants $K_k > 0$ and $H_k \geq 0$ such that $\operatorname{Ric}_{M_k}(v) \geq K_k$ and $\partial_v \mathcal{V}_{M_k} \geq -H_k$ hold for all unit vectors $v \in TM_k$, while $\lim_{k \to \infty} K_k = \infty$ and $\lim_{k \to \infty} H_k/K_k = 0$.*

(b) *There is a constant $K_k > 0$ such that $\operatorname{Ric}_{M_k}(v) + \partial_v^2 \mathcal{V}_{M_k} \geq K_k$ holds for all unit vectors $v \in TM_k$, while $\lim_{k \to \infty} K_k = \infty$.*

Then $\{(M_k, F_k, \mu_k)\}_{k \in \mathbb{N}}$ is a Lévy family.

Proof. We assume (a) and observe that Proposition 9.6(i) combined with the assumption $\operatorname{diam} M_k \geq 1$ says that

$$\alpha_{(M_k, F_k, \mu_k)}(\operatorname{diam} M_k \cdot r) \leq \alpha_{(M_k, F_k, \mu_k)}(r) \leq 2e^{2H_k^2/K_k}e^{-K_k r^2/8} = 2e^{(2H_k^2/K_k^2 - r^2/8)K_k}.$$

This completes the proof of (a), and (b) is similar. □

References

[AT]   J. C. Álvarez-Paiva and A. C. Thompson, *Volumes in normed and Finsler spaces*, A sampler of Riemann-Finsler geometry, 1–48, Math. Sci. Res. Inst. Publ., **50**, Cambridge Univ. Press, Cambridge, 2004.

[AGS]   L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures, Birkhäuser Verlag, Basel, 2005.

[Au]   L. Auslander, *On curvature in Finsler geometry*, Trans. Amer. Math. Soc. **79** (1955), 378–388.

[BE]   D. Bakry and M. Émery, *Diffusions hypercontractives* (French), Séminaire de probabilités, XIX, 1983/84, 177–206, Lecture Notes in Math., 1123, Springer, Berlin, 1985.

[BCS]   D. Bao, S.-S. Chern and Z. Shen, An introduction to Riemann-Finsler geometry, Springer-Verlag, New York, 2000.

[BB1]   P. Bernard and B. Buffoni, *The Monge problem for supercritical Mañé potentials on compact manifolds*, Adv. Math. **207** (2006), 691–706.

[BB2]   P. Bernard and B. Buffoni, *Optimal mass transportation and Mather theory*, J. Eur. Math. Soc. (JEMS) **9** (2007), 85–121.

[Be]   J. Bertrand, *Existence and uniqueness of optimal maps on Alexandrov spaces*, Adv. Math. **219** (2008), 838–851.

[Br]   Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), 375–417.

[Ch]   I. Chavel, Riemannian geometry. A modern introduction. Second edition, Cambridge University Press, Cambridge, 2006.

[CMS1]   D. Cordero-Erausquin, R. J. McCann and M. Schmuckenschläger, *A Riemannian interpolation inequality á la Borell, Brascamp and Lieb*, Invent. Math. **146** (2001), 219–257.

[CMS2]   D. Cordero-Erausquin, R. J. McCann and M. Schmuckenschläger, *Prékopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport*, Ann. Fac. Sci. Toulouse Math. (6) **15** (2006), 613–635.

[FF]   A. Fathi and A. Figalli, *Optimal transportation on non-compact manifolds*, to appear in Israel J. Math.

[FV]   A. Figalli and C. Villani, *Strong displacement convexity on Riemannian manifolds*, Math. Z. **257** (2007), 251–259.

[Ga]   R. J. Gardner, *The Brunn-Minkowski inequality*, Bull. Amer. Math. Soc. (N.S.) **39** (2002), 355–405.

[GM]   M. Gromov and V. D. Milman, *A topological application of the isoperimetric inequality*, Amer. J. Math. **105** (1983), 843–854.

[JKO]   R. Jordan, D. Kinderlehrer and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), 1–17.

[Le]   M. Ledoux, The concentration of measure phenomenon, American Mathematical Society, Providence, RI, 2001.

[Lo]   J. Lott, *Some geometric properties of the Bakry-Émery-Ricci tensor*, Comment. Math. Helv. **78** (2003), 865–883.

[LV1]   J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, to appear in Ann. of Math.

[LV2]   J. Lott and C. Villani, *Weak curvature conditions and functional inequalities*, J. Funct. Anal. **245** (2007), 311–333.

[Mc]   R. J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. **11** (2001), 589–608.

[Oh1]   S. Ohta, *On the measure contraction property of metric measure spaces*, Comment. Math. Helv. **82** (2007), 805–828.

[Oh2]   S. Ohta, *Products, cones, and suspensions of spaces with the measure contraction property*, J. Lond. Math. Soc. (2) **76** (2007), 225–236.

[Oh3]   S. Ohta, *Gradient flows on Wasserstein spaces over compact Alexandrov spaces*, to appear in Amer. J. Math.

[Oh4]   S. Ohta, *Uniform convexity and smoothness, and their applications in Finsler geometry*, to appear in Math. Ann.

[OS]   S. Ohta and K.-T. Sturm, *Heat flow on Finsler manifolds*, to appear in Comm. Pure Appl. Math.

[Ol1]   Y. Ollivier, *Ricci curvature of metric spaces*, C. R. Math. Acad. Sci. Paris **345** (2007), 643–646.

[Ol2]   Y. Ollivier, *Ricci curvature of Markov chains on metric spaces*, J. Funct. Anal. **256** (2009), 810–864.

[OV]   F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), 361–400.

[Qi]   Z. Qian, *Estimates for weighted volumes and applications*, Quart. J. Math. Oxford Ser. (2) **48** (1997), 235–242.

[RR]   S. T. Rachev and L. Rüschendorf, Mass transportation problems. Vol. I, Springer-Verlag, New York, 1998.

[vR]   M.-K. von Renesse, *On local Poincaré via transportation*, Math. Z. **259** (2008), 21–31.

[vRS]   M.-K. von Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Comm. Pure Appl. Math. **58** (2005), 1–18.

[Sa]   G. Savaré, *Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds*, C. R. Math. Acad. Sci. Paris **345** (2007), 151–154.

[Sh1]   Z. Shen, *Volume comparison and its applications in Riemann-Finsler geometry*, Adv. Math. **128** (1997), 306–328.

[Sh2]   Z. Shen, Lectures on Finsler geometry, World Scientific Publishing Co., Singapore, 2001.

[St1]   K.-T. Sturm, *Convex functionals of probability measures and nonlinear diffusions on manifolds*, J. Math. Pures Appl. **84** (2005), 149–168.

[St2]   K.-T. Sturm, *On the geometry of metric measure spaces*, Acta Math. **196** (2006), 65–131.

[St3]   K.-T. Sturm, *On the geometry of metric measure spaces.* II, Acta Math. **196** (2006), 133–177.

[Vi1]   C. Villani, Topics in optimal transportation, American Mathematical Society, Providence, RI, 2003.

[Vi2]   C. Villani, Optimal transport, old and new, Springer-Verlag, Berlin, 2009.